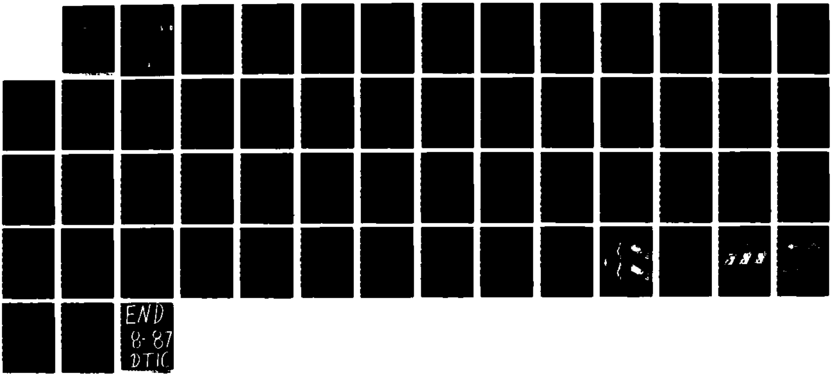
AD-A181 903 SMOOTHNESS PRIORS IN TIME SERIES(U) STANFORD UNIV CA DEPT OF STATISTICS W GERSCH ET AL 02 JUN 87 TR-391 N00014-86-K-0156 1/1

UNCLASSIFIED F/G 12/3 NL



MICROCOPY RESOLUTION TEST CHART

12

SMOOTHNESS PRIORS IN TIME SERIES

BY

WILL GERSCH and GENSHIRO KITAGAWA

TECHNICAL REPORT NO. 391
JUNE 2, 1987

PREPARED UNDER CONTRACTS
N00014-86-K-0156 and N00014-83-K-0238
(NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

SMOOTHNESS PRIORS IN TIME SERIES

BY

WILL GERSCH and GENSHIRO KITAGAWA

TECHNICAL REPORT NO. 391

JUNE 2, 1987

Prepared Under Contracts
N00014-86-K-0156 and N00014-83-K-0238
(NR-042-267)
For the Office of Naval Research
Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. INTRODUCTION 1
2. SMOOTHNESS PRIORS MODELING,LEAST SQUARES ANALYSIS 5
2.1 SMOOTHNESS PRIORS BAYESIAN MODELING 5
2.2 SMOOTHNESS PRIORS IN THE FRQEUENCY DOMAIN 9
2.3 DISCUSSION 14
3. STATE SPACE GAUSSIAN SMOOTHNESS PRIORS MODELING 16
3.1 NONSTATIONARY MEAN SMOOTHNESS PRIORS STATE SPACE MODELING 16
3.2 RECURSIVE ESTIMATION OF STATE AND LIKELIHOOD COMPUTATION 19
3.3 NONSTATIONARY COVARIANCE MODELING 21
3.4 EXAMPLES 25
3.5 COMMENTS, DISCUSSION 28
4. STATE SPACE NON GAUSSIAN MODELING OF NONSTATIONARY TIME SERIES 31
4.1 THE NON GAUSSIAN STATE SPACE MODEL 32
4.2 NUMERICAL EXAMPLES 34
4.3 COMMENTS, DISCUSSION 36
5. SUMMARY 38
REFERENCES 39

1. INTRODUCTION

Several different kinds of stationary and nonstationary time series modeling problems are considered here from a Bayesian-smoothness priors approach. The smoothness priors specify the prior distribution of the time series model parameters.

The term "smoothness priors" is very likely due to Shiller (1973). Shiller modeled the distributed lag (impulse response) relationship between the input and output of economic time series under difference equation "smoothness" constraints on the distributed lags. A tradeoff of the goodness-of-fit of the solution to the data and the goodness-of-fit of the solution to a smoothness constraint was determined by a single smoothness tradeoff parameter. Shiller did not offer an objective method of choosing the smoothness tradeoff parameter. Akaike, (1980), completed the analysis initiated by Shiller. Akaike developed and exploited the concept of the likelihood of the Bayesian model and used a maximization of the likelihood procedure for determining the smoothness tradeoff parameter. (In Bayesian terminology, the smoothness tradeoff parameter is referred to as the "hyperparameter", Lindley and Smith, 1972.) The origin of Shiller-Akaike smoothness priors can be seen in a smoothing problem posed by Whittaker (1923). The smoothing problem context is now understood to be common to a variety of other statistical data analysis problems including density estimation and image analysis (Titterington 1985).

In the problem treated by Whittaker, the observations $y_n, n=1,\dots,N$ are given. They are assumed to consist of the sum of a "smooth" function and observation noise,

$$y_n = f_n + \epsilon_n. \tag{1.1}$$

The problem is to estimate the unknown $f_n, n=1,\dots,N$. In a time series interpretation of this problem, $f_n, n=1,\dots,N$ is the trend of a nonstationary mean time series. A typical approach to this problem is to use a class of parametric models. The quality of the analysis is completely dependent upon the appropriateness of the assumed model class. A flexible model is desirable. In this context, Whittaker suggested that the solution balance a tradeoff of goodness of fit to the data and goodness of fit to a smoothness criterion. This idea was realised by minimizing

$$[\sum_{n=1}^{N} (y_n - f_n)^2 + \mu^2 \sum_{n=1}^{N} (\nabla^k f_n)^2] \tag{1.2}$$

for some appropriately chosen smoothness tradeoff parameter μ^2. In (1.2) $\nabla^k f_n$ expresses a kth-order difference constraint on the solution f, with $\nabla f_n = f_n - f_{n-1}$, $\nabla f_n^2 = \nabla(\nabla f_n)$, etc. (Whittaker's original solution was not expressed in a Bayesian context. Whittaker and Robinson (1924) does invoke a Bayesian interpretation of this problem.)

The properties of the solution to the problem in (1.1)-(1.2) are clear. If $\mu^2=0$, $f_n = y_n$ and the solution is a replica of the observations. As μ^2 becomes increasingly large, the smoothness constraint dominates the solution and the solution satisifies a kth order constraint. For large μ^2 and $k=1$, the solution is a constant, for $k=2$, it is a straight line etc.. Whittaker left the choice of μ^2 to the investigator.

Kohn and Ansley (1987) demonstrate that the signal extraction problem of (1.1) and the smoothing problem of (1.2) are equivalent problem statements. The equivalence also holds for broad variations of signal extraction and smoothing problems. All of the time series analysis problems that we treat here are variations of the signal extraction/smoothing problem in (1.1) and (1.2).

An implication of Akaike (1980) is that a Bayesian interpretation of the smoothing problem in (1.2) implies that the difference equation constraint is a stochastically perturbed sero-mean unknown variance difference equation. The stochastically perturbed difference equation constraint in the trend estimation problem is a smoothness priors constraint in the time domain. Akaike (1980), considered other time domain smoothness priors constraint problems including the Shiller distributed lag problem and the seasonal adjustment of time series. Ishiguro et al. (1981) used time domain smoothness priors constraints and fixed effects regression in an analysis of tidal effects. Akaike (1979) employed a frequency domain smoothness priors constraint on the distributed lag parameters in the Shiller problem. Gersch and Kitagawa (1984) and Kitagawa and Gersch (1985a) are other frequency domain smoothness priors time series problem analyses.

Shiller (1973), Akaike (1980), and all of the aforementioned smoothness priors analyses, are Bayesian analyses of the linear model with Gaussian stochastic constraints and Gaussian disturbances. The critical ideas in smoothness priors are the likelihood of the Bayesian model and the use of likelihood as a measure of the goodness of fit of the model. In our analysis, hyperparameters have interpretations as noise to signal ratios and they have a remarkable role in the analysis. The maximisation of the likelihood of a small number of hyperparameters permits the robust modeling of a time series with relatively complex structure and a very large number of implicitly inferred parameters. When we consider alterntative smoothness priors models, with different distributional assumptions or different numbers of parameters to model the same data, we use the Akaike AIC statistic (Akaike 1973), to choose between candidate models. Kitagawa (1987) is a smoothness priors state space modeling of nonstationary time series in which neither the system noise or the observation noise are necessarily Gaussian distributed.

The original Whittaker problem has also given rise to work on splines in numerical analysis and to related smoothing problem analysis, particularly by Wahba (1977),(1982). Smoothness priors relates to the ill-posed problems and problems of statistical regularisation that have been considered extensively in the Soviet Union by Tikhonov (1965) and his associates. Also related are the "bump hunting"-penalised likelihood methods, Good and Gaskins (1980) and Wecker and Ansley (1983) and O'Sullivan et al. (1986). Vigorous work, primarily at the Institute of Statistical Mathematics, Tokyo, has resulted in the application of smoothness priors methods to a variety of applications, other than the ones we discuss here. These applications include the seasonal adjustment of time series, (Akaike 1980b), tidal analysis (Ishiguro et al. 1981), binary regression (Ishiguro and Sakamoto 1983), cohort analysis (Nakamura 1986), and density estimation (Tanabe and Sagae 1987).

Smoothness priors problems that are amenable to analysis by least squares algorithms are treated in Section 2. The likelihood of the Bayesian model, as done by Akaike, is in Section 2.1. The Bayesian solution to the smoothing problem originally posed by Whittaker is also shown there. In that problem, the smoothness priors constraints are time domain constraints. The priors are expressed as zero-mean unknown variance stochastically perturbed kth-order random walk difference equations. In Section 2.2, the estimation of the power spectral density from short duration stationary time series illustrates the use of frequency domain smoothness priors constraints. In Section 3 several smoothness priors nonstationary time series problems, which are amenable to a Kalman filter state space method of analysis, including examples of the modeling of nonstationary mean and nonstationary covariance time series, are shown. All of the aforementioned treat a linear model, Gaussians distributions situation. That method is generalized in Section 4. There we show a smoothness priors state space not necessarily Gaussian nonstationary time series analysis method. Finally, Section 5 is a summary.

2 SMOOTHNESS PRIORS MODELING: LEAST SQUARES ANALYSIS

In Section(2.1) we review the concept of smoothness priors Bayesian modeling as introduced in Akaike (1980). That method is applied to the problem addressed by Whittaker (1923), the estimation of a trend in white noise. The smoothness priors constraint is expressed as a kth order random walk with a zero-mean, unknown variance perturbation. The variance is a hyperparameter of the prior distribution. The constraint is a time domain constraint on the priors. In Section 2.2 we introduce the notion of frequency domain constraint on the priors. That method is used in the estimation of the power spectral density of a stationary time series. Section 2.3 is a discussion.

The frequency domain smoothness priors method is particularly suited for the situation in which only a short span of data is available for analysis. In that case, the results of conventional parametric model analysis methods are particularly sensitive to the choice of model order. We circumvent that problem, using the frequency domain smoothness priors, by tending to fit models that are "too long". Those priors reflect the integrated squared kth derivative with respect to frequency of the departure from model smoothness. The estimation of the model parameters and an additional small number of hyperparameters is required. The maximization of the likelihood of the hyperparameters is the critical computation.

2.1 SMOOTHNESS PRIORS BAYESIAN MODELING

Consider the linear regression model subject to Bayesian-stochastic constraints

$$\begin{pmatrix} y \\ \theta \end{pmatrix} \sim N\left(\begin{bmatrix} X\theta \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 I & 0 \\ 0 & \lambda^2 D^{-1} D^{-T} \end{bmatrix} \right). \tag{2.1.1}$$

The dimensions of the matrices in (2.1.1) are y: nx1; X: nxp; θ: px1. σ^2 and λ^2 are unknown. y is the vector of observed data, X and D are assumed known. θ is the normally distributed prior parameter vector. The observation noise variance is σ^2. In this conjugate family Bayesian situation (Berger 1985), the mean of the posterior normal distribution of the parameter vector θ minimizes

$$\left|\left|y-X\theta\right|\right|^2 + \lambda^2\left|\left|D\theta\right|\right|^2. \tag{2.1.2}$$

If λ^2 were known, the computational problem in (2.1.2) could be solved by an ordinary least squares computation. The solution for $\hat{\theta}$, the posterior mean, is the minimizer of

$$\left|\left|\begin{pmatrix} y \\ 0 \end{pmatrix} - \begin{pmatrix} X \\ \lambda D \end{pmatrix} \theta\right|\right|^2 . \tag{2.1.3}$$

That solution is

$$\hat{\theta} = \left[X^TX + \lambda^2D^TD\right]^{-1}X^Ty \tag{2.1.4}$$

with the residual sum of squares,

$$SSE(\hat{\theta},\lambda^2) = y^Ty - \hat{\theta}^T\left[X^TX + \lambda^2D^TD\right]^{-1}\hat{\theta}. \tag{2.1.5}$$

For a smoothness priors interpretation of the problem in (2.1.1) and (2.1.2), multiply (2.1.2) by $-1/2\sigma^2$ and exponentiate. Then the θ that minimizes (2.1.2) also maximizes

$$\exp\left\{-\frac{1}{2\sigma^2}\left|\left|y - X\theta\right|\right|^2\right\}\exp\left\{-\frac{\lambda^2}{2\sigma^2}\left|\left|D\theta\right|\right|^2\right\} . \tag{2.1.6}$$

In (2.1.6), the posterior distribution interpretation of the parameter vector θ is that it is proportional to the product of the conditional data distribution (likelihood), $p(y|\, X,\theta,\sigma^2)$, and a prior distribution, $\pi(\theta|\, \lambda^2,\sigma^2)$ on θ,

$$\pi(\theta|\, y,\lambda^2,\sigma^2) \;\alpha\; p(y|\, X,\theta,\sigma^2)\pi(\theta|\, \lambda^2,\sigma^2) . \tag{2.1.7}$$

The integration of (2.1.7) yields $L(\lambda^2,\sigma^2)$, the likelihood for the unknown parameters λ^2 and σ^2,

$$L(\lambda^2,\sigma^2) = \int_{-\infty}^{\infty} \pi(\theta|\, y,\lambda^2,\sigma^2)d\theta . \tag{2.1.8}$$

I.J. Good (1965) referred to the maximization of (2.1.8) as a Type II maximum likelihood method. Since $\pi(\theta|\, y,\lambda^2,\sigma^2)$ is normally distributed, (2.1.8) can be expressed in the closed form, (Akaike 1980),

$$L(\lambda^2,\sigma^2) = (2\pi\sigma^2)^{-N/2}|\,\lambda^2 D^T D|^{1/2}|\,X^T X + \lambda^2 D^T D|^{-1/2}\exp\left\{\frac{-1}{2\sigma^2} SSE(\hat{\theta},\lambda^2)\right\} . \tag{2.1.9}$$

The maximum likelihood estimator of σ^2 is

$$\hat{\sigma}^2 = SSE(\hat{\theta},\lambda^2)/N . \tag{2.1.10}$$

It is convenient to work with -2 log likelihood. Using (2.1.10) in (2.1.9) yields

$$-2logL(\lambda^2,\hat{\sigma}^2) = Nlog2\pi + NlogSSE((\hat{\theta}^2,\lambda^2)/N) + \log|\,X^T X + \lambda^2 D^T D| - \log|\,\lambda^2 D^T D| + N. \tag{2.1.11}$$

A practical way to determine the value of λ^2 for which the -2log-likelihood is minimized, is to compute the likelihood for discrete values of λ^2 and search the discrete -2log likelihood-hyperparameter space for the minimum. Akaike (1980) is very likely the first practical use of the likelihood of the Bayesian model and the use of the likelihood of the hyperparameters, as a measure of the goodness of fit of a model to data.

Estimating a Trend

Here we return to the original problem posed in the Introduction. We use the notation $f_n = t_n$ where t_n is the trend at time n to emphasize the fact that we are estimating the mean of a nonstationary mean time series. A critically important observation is that from the stochastic regression or Bayesian point of view, the difference equation constraints in the Whittaker problem are stochastic. That is, $\nabla^k t_n = w_n$, with w_n assumed to be a normally distributed zero-mean sequence with unknown variance τ^2 . For example for $k=1$ and $k=2$ those constraints are:

$$t_n = t_{n-1} + w_n; \tag{2.1.12}$$

$$t_n = 2t_{n-1} - t_{n-2} + w_n .$$

A parameterization which relates the trend estimation problem to the earlier development in this section is $\tau^2 = \sigma^2/\lambda^2$. Corresponding to the matrix D in (2.1.2), for $k=1$ and $k=2$, the smoothness constraints can be expressed in terms of the following NxN constraint matrices:

$$D_1 = \begin{bmatrix} \alpha & & & & & \\ -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \cdot & & & \\ & 0 & & \cdot\ \cdot & & \\ & & & \cdot\ \cdot & \cdot & - \\ & & & & -1 & 1 \end{bmatrix}, \quad D_2 = \begin{bmatrix} \alpha & & & & & & \\ -\beta & \beta & & & & & \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & 0 & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & & \cdot \\ & & & & 1 & -2 & 1 \end{bmatrix}. \tag{2.1.13}$$

In (2.1.13) α and β are small numbers that are chosen to satisfy initial conditions.

For fixed k and fixed λ^2 the least squares solution can be simply expressed in the form of (2.1.13). For example with $k = 2$, the solution $\{t_n, n = 1, \ldots, N\}$ satisfies

$$\left\| \begin{pmatrix} y \\ 0 \end{pmatrix} - \begin{pmatrix} I \\ \lambda D \end{pmatrix} t \right\|^2 . \tag{2.1.14}$$

Note that the problem in (2.1.14) is a version of the Bayesian linear stochastaic regression problem in (2.1.3) with $\theta = t = (t_1, \ldots, t_N)^T$, $X = I$, the NxN identity matrix, and $D = D_1$ or D_2 . From (2.1.3),the solution to (2.1.14), with $D = D_2$, is

$$\hat{t} = [I + \lambda^2 D_2^T D_2]^{-1} y \tag{2.1.15}$$

and the value of $SSE(\hat{\theta}, \lambda^2)$ is given by (2.1.5) with $\hat{\theta} = \hat{t}, X = I, D = D_2$. The smoothing problem expression of (2.2.15) is that the solution vector is: $\hat{t} = (t_{1|N}, \ldots, t_{N|N})^T$. The least squares problem in (2.1.14), with $D = D_2$, is solved for discrete values of λ^2D and the -2 log likelihood-hyperparameter space is searched for a minimum. From (2.1.11), the minimized value of -2log likelihood for this problem is:

$$-2logL(\hat{\lambda}^2, \hat{\sigma}^2) = Nlog2\pi + NlogSSE((\hat{t}^2, \hat{\lambda}^2)/N)) + log|\,\hat{\lambda}^2 D_2^T D_2 + I| - log|\,\hat{\lambda}^2 D_2^T D_2| + N. \tag{2.1.16}$$

The numerical values of SSE($\hat{t}^2, \lambda^2$) and of the determinants in (2.1.16) are transparent in a least squares algorithm analysis of (2.1.14). Since $\lambda = \sigma/\tau$, λ^2 has a noise-to-signal ratio interpretation. Smaller λ corresponds to smoother trends.

An Example of Trend Estimation

We consider the example of an asymetrically truncated normal density-like function in the presence of additive noise, $N(0,\sigma^2)$. Figure 1a shows the smooth function $t_n, n=1,\ldots,N$ and the superposition of t_n and the additive noise. The problem is: Given the noisy observations y_n, estimate the unknown smooth function that is in the noise, i.e. specify $t_{n|N}, n=1,\ldots,N$. We solved the least squares computational problem in (2.1.14) using the Householder transformation method. -2 log likelihood of the hyperparameter model is computed from (2.1.16)

The critical role of the hyperparameter is transparent in this example Figures 1b,c,d show the estimated trend for values of the hyperparameters that are too small, (λ^2 = 0.00001), and too large (λ = 0.1) as well as the hyperparmeter for which -2log likelihood is minimised (λ = 0.00136). As anticipated, with the hyperparameter defined as indicated above, the estimated trend for a too large value of the hyperparameter is too bumpy and the estimated trend for a too small value of the hyperparameter is too smooth.

It is important to note that in this example, although the truncated Gaussian satisfies $\nabla^2 \log t_n = 0$, we estimate the trend with the "incorrect" model $\nabla^2 t_n = w_n$, the stochastically perturbed second order difference equation. The point is that a priori we do not know a correct expression for the underlying smooth function in (1.1). Different hyperparameter values result in solutions of the stochastically perturbed second order difference equation with very different smoothness properties. The best of those solutions yields a very good approximation to the original unknown smooth function. This key observation was referred to by Shiller, (1973), as the "flexible ruler approach".

2.2 SMOOTHNESS PRIORS IN THE FREQUENCY DOMAIN

The smoothness priors in the estimation of the mean value of a nonstationary time series was expressed as a time domain, stochastically perturbed difference equation constraint on the evolution of the trend. Smoothness priors contraints can also be expressed in the frequency domain In this section, we illustrate the use of frequency domain priors for the estimation of the power spectral density of a stationary time series.

A Long AR Model For Spectral Estimation

A smoothness priors-long autoregressive (AR) model approach is used here for spectral density estimation.

The classical windowed periodogram method of spectral estimation is satisfactory for spectral analysis when the data set is "long". The alternative of spectral estimation via the fitting of parametric models to moderate length data spans became popular in the last decade, Kesler(1986). When the data span is relatively short, three facts render parametric modeling methods of spectral estimation statistically unreliable. One is the instability or small sample variability of whatever statistic is used for determining the best order of parameteric model fitted to the data. The second is that usually the "parsimonious" parametric model is not a very good replica of the system that generated the data. The third is that the spectral density of the fitted parametric model can not possibly be correct. Independent of which parametric model order is selected, there is information in the data to select models of different orders. A Bayesian estimate of power spectral density requires that the spectral density of parametric models of different model orders be weighted in accordance with the likelihood and the prior of the model order of different models.

The smoothness priors AR model of spectral estimation alleviates this problem. A particular class of frequency domain smoothness priors is assumed for the coefficients of AR model order M, with M relatively large. The likelihood of the hyperparameters that characterize the class of smoothness priors is maximized to yield the best AR model of order M with the best data dependent priors. (A more complete treatment of the modeling discussed here is in Kitagawa and Gersch, 1985a.)

The Smoothness Priors Long AR Model

Consider the autoregressive model of order M,

$$y_n = \sum_{m=1}^{M} a_m y_{n-m} + \epsilon_n \tag{2.2.1}$$

In (2.2.1) $\{\epsilon_n\}$ is a Gaussian white noise with mean zero and variance σ^2. A least squares fit of the AR model to the data, $y_1,...,y_N$, with the first M observations $y_{1-M}, y_{2-M},...,y_0$ treated as given constants, leads to the minimization of

$$\sum_{n=1}^{N} |y_n - \sum_{m=1}^{M} a_m y_{n-m}|^2 \tag{2.2.2}$$

If M is comparable to N, the result of the least squares computation can be meaningless. The smoothness priors solution mitigates this difficulty by considering the solution of the constrained least squares problem. We consider a frequency domain smoothness priors constraint on the distribution of the AR model parameters. The frequency response function of the whitening filter of the AR process given by

$$A(f) = 1 - \sum_{m=1}^{M} a_m \exp[-2\pi imf]. \tag{2.2.3}$$

Let a measure of the smoothness of the frequency response function be

$$R_k = \int_{-1/2}^{1/2} |\frac{d^k A(f)}{df^k}|^2 df = (2\pi)^{2k} \sum_{m=1}^{M} m^{2k} a_m^2. \tag{2.2.4}$$

From the definition in (2.2.4), a large value of R_k means an unsmooth (in the sense of differential) frequency response function. We also use the zero derivative smoothness constraint,

$$R_0 = \int_{-1/2}^{1/2} |A(f)|^2 df = 1 + \sum_{m=1}^{M} a_m^2 \tag{2.2.5}$$

as a penalty to the whitening filter.

With these constraints, and with λ^2 and ν^2 fixed, the AR model coefficients $\{a_m, m=1,...,M\}$, minimize

$$\sum_{n=1}^{N}[y_n - \sum_{m=1}^{M} a_m y_{n-m}]^2 + \lambda^2 \sum_{m=1}^{M} m^{2k} a_m^2 + \nu^2 \sum_{m=1}^{M} a_m^2. \qquad (2.2.6)$$

In (2.2.6), λ^2 and ν^2 are the hyperparameters. By a proper choice of these parameters, our estimates of the AR model coefficients balance the tradeoff between the infidelity of the model to the data and the infidelity of the model to the frequency domain smoothness constraints. For completeness, to within a constant, the Gaussian priors on the AR model coefficients corresponding to the R_0 and R_k constraints are

$$\exp-\frac{\lambda^2}{2}\sigma^2 \sum_{m=1}^{M} m^{2k} a_m^2 \exp-\frac{\nu^2}{2}\sigma^2 \sum_{m=1}^{M} a_m^2. \qquad (2.2.7)$$

Following our earlier discussion, define the matrices D and a and the matrices X and y by

$$D = \begin{bmatrix} (\nu^2+\lambda^2)^{1/2} & & & \\ & (\nu^2+2^{2k}\lambda^2)^{1/2} & & \\ & & \ddots & \\ & & & (\nu^2+M^{2k}\lambda^2)^{1/2} \end{bmatrix}, a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix}, \quad X = \begin{bmatrix} y_0 & \cdots & y_{1-M} \\ y_1 & \cdots & y_{2-M} \\ \vdots & & \vdots \\ y_{N-1} & \cdots & y_{N-M} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad (2.2.8)$$

Then the AR model coefficients satisfy

$$\hat{a} = (X^T X + D^T D)^{-1} X^T y, \qquad (2.2.9)$$

and the residual sum of squares is

$$S(\lambda^2,\nu^2) = y^T y - \hat{a}^T (X^T X + D^T D)\hat{a}. \qquad (2.2.10)$$

The likelihood of the hyperparamter model is

$$L(\lambda^2,\nu^2,\sigma^2) = (\frac{1}{2\pi\hat{\sigma}^2})^{N/2} | D^T D |^{1/2} | X^T X + D^T D |^{-1/2} \exp\left\{\frac{-1}{2\sigma^2} S(\lambda^2,\nu^2)\right\}. \qquad (2.2.11)$$

Given λ^2 and ν^2, the maximum likelihood estimate of σ^2 is, $\hat{\sigma}^2 = S(\lambda^2,\nu^2)/N$. The ML estimates of λ^2 and ν^2 are obtained by minimizing

$$-2logL(y| \lambda^2,\nu^2,\sigma^2) = N\log 2\pi\sigma^2 + \log| D^T D| - log| X^T X + D^T D| + N. \qquad (2.2.12)$$

with respect to λ^2 and ν^2. Computation of the likelihood over a discrete k,λ^2,ν^2 parameter grid and searching over the resulting discrete likelihood-hyperparameter space for the minimum of -2 log likelihood yields the desired smoothness priors long AR model.

The frequency domain smoothness priors constraint used here has an interpretation as a constraint on the smoothness of the whitening filter of the AR model. (The 0th derivative has an energy constraint interpretation.) An important facet of our computations is that they are computationally tractable. That allows us to remain within the framework of the general linear model.

An Example, Analysis of Canadian Lynx Data

The data example discussed here is the analysis of the Canadian Lynx data (n=114). Other examples are shown in Kitagawa and Gersch (1985a).

The AR model order was set to 20 and up to the fourth order smoothness prior constraint was tried. The hyperparameters λ and ν were searched over the discrete values $\lambda = 2^{j-3k}\sigma_0, j=1,...,10$ where σ_0^2 is the sample variance of the data and $\nu = i^2\sigma_0$, $i=0,1,...,4$, for each value of the order of the smoothness prior constraint.

The overall best model was $k=1, \nu = 0, \lambda = 0.173$. The Bayesian estimate of the spectrum is shown in Figure 2a. For comparison, AR models of order up to 20 were fitted by the least squares method. The AIC best order was 11. The AIC criterion-AR modeled spectrum is shown in Figure 2b. In the Bayesian model spectral estimate, the peaks at the high frequencies are significantly reduced compared to the AR model model spectrum estimate, while the ones in the lower frequencies remain unchanged. Figure 2c shows the superimposed estimated spectra obtained from AR models with different model orders. From Figures 2b,2c we see that the shape of the two rightmost peaks of Figure 2.1B vary considerably with model order. Thus they are not estimated reliably by fixed order models. That is typical of the problem of estimating spectral density by fixed order parametric models. If the model fitted to the data is not in the class of models which generated the data, the model fitting is only approximate. The selection of the best non Bayesian parametric model ignores the evidence, in the Bayesian sense, for other parametric models when in fact it should be taken into account. The suppression of those peaks by the smoothness priors-long AR model method, shown in Figure 2a, therefore seems quite reasonable.

2.3 DISCUSSION

The variation of the behavior of the solution, from very rough to very smooth, in the trend estimation problem under the smoothness priors constraints for different values of the hyperparameters, is characteristic of the profound effect of the hyperparameter. The log likelihood of the hyperparameter versus the hyperparameter changes gradually in the vicinity of the maximum log likelihood. That fact permits a discrete likelihood-hyperparameter search procedure to be used in conjunction with a Householder transformation algorithm to realise a reasonable computational procedure.

The stochastically perturbed difference equation constraint in the trend estimation problem is a time domain smoothness priors constraint. Akaike (1980), considered other time domain smoothness constraint problems including the Shiller distributed lag problem and the seasonal adjustment of time series. Ishiguro et al. (1981), used time domain smoothness constraints and fixed effects regression in the analysis of earth tide data. The Householder transformation least squares algorithm FORTRAN programs BAYSEA and BAYTAP-G in TIMSAC-84 (Akaike et al. 1985), are suitable for seasonal adjustment and tidal analyses respectively.

Akaike (1979) illustrated a frequency domain smoothness prior for the solution of the Shiller (1973) impulse response estimation problem. We used frequency domain smoothness priors here for spectral density estimation. Gersch and Kitagawa (1984) is an application of the frequency domain-smooothness priors method to transfer function estimation. Our smoothness priors method is particularly suited for the situation in which only a short span of data is available for analysis. In that situation, the results of conventional parametric model analysis methods are particularly sensitive to the choice of model order. We circumvent that problem, with the Bayesian smoothness priors method, by tending to fit models that are "too long". The model parameters are specified as the solution to a constrained least squares problem in which the constraints are expressed in the frequency domain. The likelihood of the hyperparameters is readily computable in a least squares framework with the frequency domain priors.

The goodness of the choice of the fequency domain smoothness priors can be appraised by evaluating its performance in various conceivable situations. The smoothness priors-long AR model gives reasonable results in the analysis of the real Lynx data in comparison with the minimum AIC-AR model method. Kitagawa and Gersch (1985a) show smoothness priors- long AR model results that were superior to the minimum AIC-AR model method in a simulated-two sine waves in noise case, when the data actually corresponds to an ARMA model. This flexibility of performance is what is desired from a Bayesian model.

Also in Kitagawa and Gersch (1985a), a Monte Carlo study of the expected entropy experiment was done to appraise the performance of the smoothness priors-long AR model for spectral estimation against performance of parametric AR models whose order was determined by Akaike's AIC criterion. The smoothness priors-long AR model method was superior to the minimum AIC-AR model method in the two simulation model cases studied. In one case, the simulation model was in the AR model class. In the other case, the simulation model was not in the AR model class. Thus, the example shown here and the Monte Carlo study reported in Kitagwa and Gersch (1985a) are evidence to support the soundness of our empirical frequency domain smoothness priors approach.

3 STATE SPACE GAUSSIAN SMOOTHNESS PRIORS MODELING

A state space modeling approach for the linear model with Gaussian system and observation noise that is the equivalent of the least squares computational approach to smoothness priors modeling, was shown in Brotherton and Gersch (1981) and Kitagawa (1981). The state space smoothness priors modeling method was applied to the modeling of nonstationary mean and nonstationary covariance time series, Gersch and Kitagawa (1983a,1985) and Kitagawa and Gersch (1984,1985b).

In the modeling of nonstationary time series discussed below, there tends to be more parameters than data. In that case, attempts to fit the parameters by least squares or any other ordinary means will yield poor parameter estimates. The smoothness priors permit the model parameters to be expressed implicitly as the solution of zero-mean unknown variance stochastically perturbed difference equations. The variances are hyperparameters of the prior distribution of the parameters. One interpretation of the role of the smoothness priors is that they permit a realization of a computational procedure to estimate the model parameters.

In this section, computational procedures for the modeling of nonstationary mean and nonstationary covariance time series are discussed that are variations of the procedures discussed in our previous papers. Examples are shown in Section 3.4. A discussion of other problems treated by the smoothness priors-state space-linear-Gaussian model and comments appear in Section 3.5. Kalman filter, prediction and smoothing formulas and computation of the likelihood of the linear Gaussian model are shown in Section 3.2.

3.1 NONSTATIONARY MEAN SMOOTHNESS PRIORS STATE SPACE MODELING

Time series with trend and seasonal components occur for example in meteorological, oceanographic and econometric studies. Here we consider a complex nonstationary mean time series problem motivated by economic time series considerations. The economic time series nonstationary mean can be decomposed into a trend t_n, a globally stationary component v_n, a seasonal component s_n, a trading day factor d_n and an observation noise component ϵ_n,

$$y_n = t_n + s_n + v_n + d_n + \epsilon_n. \tag{3.1.1}$$

Each of the aforementioned components can be modeled as a stochastically perturbed difference equation. The generic state space model for this decomposition can be expressed by

$$\begin{aligned} x_n &= Fx_{n-1} + Gw_n \\ y_n &= H_n x_n + \epsilon_n \end{aligned} \tag{3.1.2}$$

where F, G and H_n are MxM, MxL and $1xM$ matrices respectively. w_n and ϵ_n are each assumed to be zero mean independent normally distributed random variables. x_n is the state vector at time n and y_n is the observation at time n. For any particular model of the time series, the matrices F, G and H_n are known and the observations are generated recursively starting from an initial state that is assumed to be normally distributed with mean x_0 and covariance matrix V_0.

The state space model that includes the local polynomial trend, stationary AR coefficient, trading day effects and observation error components can be written in the orthogonal decomposition form

$$\begin{aligned} x_n &= \begin{bmatrix} F_1 & 0 & 0 & 0 \\ 0 & F_2 & 0 & 0 \\ 0 & 0 & F_3 & 0 \\ 0 & 0 & 0 & F_4 \end{bmatrix} x_{n-1} + \begin{bmatrix} G_1 & 0 & 0 & 0 \\ 0 & G_2 & 0 & 0 \\ 0 & 0 & G_3 & 0 \\ 0 & 0 & 0 & G_4 \end{bmatrix} w_n \\ y_n &= [H_1\ H_2\ H_3\ H_{4,n}] x_n + \epsilon_n. \end{aligned} \tag{3.1.3}$$

The component models (F_j, G_j, H_j) in order, $(j=1,\ldots,4)$ represent the trend, stationary AR, seasonal and trading day effects components respectively. Some of the particular trend, AR, seasonal and trading day difference equation constraints that we have employed and that have representations in the (F_j, G_j, H_j) matrices in (3.1.3) are shown below.

The trend component t_n satisfies a kth order stochastically perturbed difference equation

$$\nabla^k t_n = w_{1,n} \tag{3.1.4}$$

where $w_{1,n}$ is an i.i.d. sequence with $w_{1,n} \sim N(0,\tau_1^2)$. (See (2.1.12).)

The stationary AR component v_n is assumed to satisfy an AR model of order p. That is given by

$$v_n = a_1 v_{n-1} + \cdots + a_p v_{n-p} + w_{2,n} . \tag{3.1.5}$$

In (3.1.5) $w_{2,n}$ is and i.i.d. sequence with $w_{2,n} \sim N(0,\tau_2^2)$. The seasonal component of the period L difference equation is

$$s_n = -s_{n-1} - s_{n-2} - \cdots - s_{n-L+1} + w_{3,n} . \tag{3.1.6}$$

In (3.1.6), $w_{3,n}$ is an i.i.d sequence with $w_{3,n} \sim N(0,\tau_3^2)$.

The trading day effect model is

$$d_n = \beta_{1,n} d_{1,n} + \cdots + \beta_{6,n} d_{6,n} \tag{3.1.7}$$

where $\beta_{i,n}$ denotes the trading-day effect factor and $d_{i,n}$ corresponds to the number of the ith day of the week at time n. Implicit in (3.1.7) is the constraint $\sum_{i=1}^{7} \beta_{i,n} = 0$. There is no stochastic component in (3.1.7).

For a general model including local polynomial trend, AR component trend, local seasonal component and trading day effect components, the state or system noise vector and observation nosie ϵ_n are assumed to i.i.d. with zero mean and diagonal covariance matrix

$$\begin{pmatrix} w_n \\ \epsilon_n \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} , \begin{pmatrix} \tau_1^2 & 0 & 0 & 0 \\ 0 & \tau_2^2 & 0 & 0 \\ 0 & 0 & \tau_3^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix} \right) . \tag{3.1.8}$$

An example of a state space model that incorporates each of the components with trend order 2, AR model order 2 and seasonal component with period L is,

$$x_n = \begin{bmatrix} t_n \\ t_{n-1} \\ \hline v_n \\ v_{n-1} \\ \hline s_n \\ \cdot \\ \cdot \\ \cdot \\ s_{n-L+2} \\ \hline \beta_{1,n} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{6,n} \end{bmatrix} = \left[\begin{array}{cc|cc|ccccc|ccc} 2 & -1 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & \cdot & 0 \\ 1 & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & \cdot & 0 \\ \hline 0 & 0 & a_1 & a_2 & 0 & \cdot & \cdot & \cdot & 0 & 0 & \cdot & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & \cdot & 0 \\ \hline 0 & 0 & 0 & 0 & -1 & \cdot & \cdot & \cdot & -1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 & 1 & \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & & & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & & & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & \cdot & \cdot & 1 & 0 & \cdot & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdot & 1 \end{array}\right] x_{n-1} + \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & 0 \end{array}\right] w_n \qquad (3.1.9)$$

$$y_n = \left[\begin{array}{cc|cc|cccc|ccc} 1 & 0 & 1 & 0 & 1 & \cdot & \cdot & 0 & d_{1,n} & \cdot\,\cdot\,\cdot & d_{6,n} \end{array}\right] x_n + \epsilon_n .$$

The smoothness priors problem that includes all of the components in the decomposition identified above correponds to the maximization of

(3.1.10)

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}[y_n - t_n - s_n - d_n]^2\right\} \cdot$$

$$\exp\left\{-\frac{\tau_1^2}{2\sigma^2}\sum_{n=1}^{N}[\nabla^2 t_n]^2\right\}\exp\left\{-\frac{\tau_2^2}{2\sigma^2}\sum_{n=1}^{N}[v_n - \sum_{i=1}^{2} a_i v_{n-i}]^2\right\}\exp\left\{-\frac{\tau_3^2}{2\sigma^2}\sum_{n=1}^{N}[\sum_{i=0}^{L-1} s_{n-i}]^2\right\} .$$

The first term in (3.1.10) corresponds to the conditional data distribution. The remaining terms in (3.1.10), in order, corresspond to the priors on the trend, the globally stochastic component and the seasonal component.

The role of the hyperparameters τ_1^2 and τ_3^2 as a measure of the uncertainty in the belief of the priors is clear from (3.1.10). Relatively small τ_1^2 (τ_3^2) imply relatively wiggly trend (seasonal) components. Relatively large τ_1^2 (τ_3^2) imply relatively smooth trend (seasonal) components. Correspondingly, the ratio of τ_j^2/σ^2, $j=1$ or 3, can be interpreted as signal-to-noise ratios. (The value of σ^2 in (3.1.10) is essentially estimated free of computational cost in the Kalman filter algorithm.)

3.2 RECURSIVE ESTIMATION OF STATE AND LIKELIHOOD COMPUTATION

Let a state space model be given by

$$\begin{aligned} x_n &= F_n x_{n-1} + G_n w_n \\ y_n &= H_n x_n + \epsilon_n, \end{aligned} \tag{3.2.1}$$

where $w_n \sim N(0, Q_n)$ and $\epsilon_n \sim N(0, R_n)$. Given the observations $y_{1,\ldots,}y_N$ and the initial conditions $x_{0|0}$, $V_{0|0}$, the one-step-ahead predictor and the filter are obtained from the Kalman filter algorithm:

Time Update (Prediction)

$$\begin{aligned} x_{n|n-1} &= F_n x_{n-1|n-1} \\ V_{n|n-1} &= F_n V_{n-1|n-1} F_n^T + G_n Q_n G_n^T. \end{aligned} \tag{3.2.2}$$

Observation Update (Filtering)

$$\begin{aligned} K_n &= V_{n|n-1} H_n^T [H_n V_{n|n-1} H_n^T + R_n]^{-1} \\ x_{n|n} &= x_{n|n-1} + K_n [y_n - H_n x_{n|n-1}] \\ V_{n|n} &= [\, I - K_n H_n] V_{n|n-1}. \end{aligned} \tag{3.2.3}$$

Using these estimatates, the smoothed value of the state x_n given the entire observation set, $y_1, \ldots, y_N$, is obtained by the fixed interval smoothing algorithm, (Anderson and Moore 1979),

$$\begin{aligned} A_n &= V_{n|n} F_n^T V_{n+1|n}^{-1} \\ x_{n|N} &= x_{n|n} + A_n [x_{n+1|N} - x_{n+1|n}] \\ V_{n|N} &= V_{n|n} + A_n [\, V_{n+1|N} - V_{n+1|n}] A_n^T. \end{aligned} \tag{3.2.4}$$

The state space representation and the Kalman filter yield an efficient algorithm for the likelihood of a time series model. The joint distribution of $y_1, \ldots, y_N$ is,

$$f(y_{1,\dots,}y_N) = \prod_{n=1}^{N} f(y_n | y_1, \dots, y_{n-1}), \tag{3.2.5}$$

with

$$f(y_n | y_1,\dots,y_{n-1}) = (2\pi v_n)^{-1/2}\exp\left\{\frac{-1}{2v_n}(y_n - H_n x_{n|n-1})^2\right\}, \tag{3.2.6}$$

$$v_n = H_n V_{n|n-1} H_n^T + R_n.$$

Then, the log likelihood, l, of the model is obtained by

$$l = -\frac{1}{2}\left[N\log 2\pi + \sum_{n=1}^{N}\log v_n + \sum_{n=1}^{N}\frac{-1}{2v_n}(y_n - H_n x_{n|n-1})^2\right], \tag{3.2.7}$$

The maximum likelihood estimate of the model parameters are obtained by maximizing (3.2.7) with respect to those parameters. The AIC is defined by

$$AIC = -2(maximum\ log-likelihood) + 2(number\ of\ parameters) \tag{3.2.8}$$

Alternative models for time series might be models with and without trading day effects or models with and without AR component effects. In each case, when we consider alternative models, the model with the smallest value of the AIC statistic is selected as the AIC best model.

In fitting a stationary model, we can utilise the theoretical mean and the theoretical covariance of the state vector as the initial values $x_{0|0}$ and $V_{0|0}$. In the nonstationary case we consider the initial vector, $x_{0|0}$, as an unknown parameter and estimate it by using the entire set of data. The log likelihood obtained by estimating the initial state vector is a natural estimate of the expected log likelihood of the predictive distribution (Akaike 1980b, and Gersch and Kitagawa 1983a).

3.3 NONSTATIONARY COVARIANCE MODELING

Here,time series with nonstationary covariances are modeled by a time varying autoregressive (AR) model with smoothness constraints on the AR parameters. Time varying AR coefficient models have been a topic of research for some time. For example, see Whittle (1965), Kozin

(1977) and Nicholls and Quinn (1985) and the extensive references therein, particularly to the use of random coefficients in econometric modeling. Earlier, in engineering applications, the modeling of nonstationary covariance time series was done by fitting locally stationary models, and by orthogonal polynomial expansions of AR coefficient models. Astrom and Wittenmark (1973, Theorem 5) express a time varying AR coefficient model that includes the possibility of random AR coefficients. Bohlin (1976) is an early application of the analysis of time series models with time-dependent coefficients. The concept of the likelihood of the Bayesian model or of hyperparameters or anything related to a smoothness prior do not appear in the earlier papers. Those are key concepts here. Kitagawa (1983) is a precedent to the material discussed in this section.

The problem in modeling nonstationary covariance time series is to achieve an efficient parameterization to capture the local and global statistical relationships in the time series. That objective is achieved here by imposing smoothness priors constraints in the form of stochastically perturbed difference equations on the evolution of the AR coefficients. The variances of the white *noise stcchastic perturbations are the hyperparameters of* the the AR coefficient distribution. The difference equations are imbedded into a state-space representation. A relatively large AR model order is chosen, the AR coefficients at each time instant are also smoothed using the frequency domain differential contraints on AR coefficients, as in the smoothness priors-long AR model for spectral estimation, Section 2.2. For each order of the differential constraint, ·he Kalman filter yields the likelihood of the hyperparameters. The smoothed estimate of the nonstationary innovations series variance is also computed. That is used in the computation of an instantaneous spectral density which is defined in terms of the instantaneous AR model coefficients and the innovations series variance.

The Time Varying AR Coefficient Model

A time varying AR coeficient model is given by

$$y_n = \sum_{i=1}^{m} a_{i,n} y_{n-i} + \epsilon_n . \qquad (3.3.1)$$

In (3.3.1), the coefficients $a_{i,n}$ are assumed to change "gradually" with time and ϵ_n is assumed to be a normally distributed white noise sequence with variance σ^2. Since there are mxN AR coefficients in the model in (3.3.1), an attempt to fit the parameters by least squares or any other ordinary means to the N observations $y_1, \ldots, y_N$, will yield poor parameter estimates. We consider the unknown AR coefficients to be random variables and impose stochastic constraints on those coefficients. Those constraints define a Gaussian smoothness prior distribution on the time history of the AR coefficients and on the spectrum at each time instant. A simple and useful model for a time varying AR coefficient model is obtained by the stochastically perturbed difference equation constraint model

$$\nabla^{k_1} a_{i,n} = \delta_{i,n}, \quad i=1,..,m . \qquad (3.3.2)$$

For convenience, in (3.3.2) $\delta_{i,n}$ is assumed to be a zero-mean Gaussian white noise sequence with variance τ_i^2 independent of i and n. That is, $\tau_i^2 = \tau^2$, $i=1,\ldots,m$.

The smoothness priors constraints on the AR coefficients mitigate the problem of overparamerization by permitting the AR coefficients to be expressed as the solution of the constrained least squares problem

$$\sum_{n=1}^{N} [y_n - \sum_{i=1}^{m} a_{i,n} y_{n-i}]^2 + \tau^2 \sum_{n=1}^{N}\sum_{i=1}^{m} [\nabla^{k_1} a_{i,n}]^2 + \lambda^2 \sum_{n=1}^{N}\sum_{i=1}^{m} i^{2k_2} a_{i,n}^2 + \nu^2 \sum_{n=1}^{N}\sum_{i=1}^{m} a_{i,n}^2 . \qquad (3.3.3)$$

In (3.3.3) m and k_1, k_2 are assumed known and τ^2, λ^2, ν^2 are the tradeoff parameters which balance the infidelity of the model to the data and the infidelity of the model to the smoothness constraints.

Similar to the analysis in Section (2.1), (3.3.3) yields a Bayesian interpretation of the least squares problem. Multiply (3.3.3) by $-1/2\sigma^2$ and exponentiate. Then , to within a constant term,

$$\exp\left\{-\frac{\tau^2}{2\sigma^2} \sum_{n=1}^{N}\sum_{i=1}^{m} [\nabla^{k_1} a_{i,n}]^2\right\} \exp\left\{-\frac{\lambda^2}{2\sigma^2} \sum_{i=1}^{m} i^{2k_2} a_{i,n}^2\right\} \exp\left\{-\frac{\nu^2}{2\sigma^2} \sum_{i=1}^{m} a_{i,n}^2\right\} \qquad (3.3.4)$$

expresses the product of the prior distribution on the smoothness of the spectrum and the prior distribution on the smoothenss of the AR parameters. The tradeoff parameters τ^2,λ^2,ν^2 are the hyperparameters of the prior distribution. As in the development in Section (2.1), the product of the conditional data distribution, (proportional to the leftmost term in (3.3.3)), and the prior distrtibution in (3.3.4) yields the posterior distribution for the AR parameters. As in (2.1.8), integration of the posterior distribution for the AR parameters yields the likelihood for the smoothness tradeoff parameters.

State Space Time Varying AR Coefficient Model

Define the kmx1 state vector at time n to be $x_n = (a_{1,n},\ldots,a_{m,n}, \ . \ . \ . \ , a_{1,n-k+1},\ldots,a_{m,n-k+1})^T$. The state space time varying AR coefficient model is

$$\begin{aligned} x_n &= Fx_{n-1} + Gw_n \\ y_n &= H_n x_n + v_n . \end{aligned} \qquad (3.3.5)$$

In (3.3.5), H_n, is a $(m+1)$xkm observation matrix, w_n is the m vector, $w_n = (\delta_{1,n}, \ . \ . \ . \ ,\delta_{m,n})^T$ and the $m+1$ vector $v_n=(\epsilon_n,\xi_n)^T$, is defined in (3.3.7). For the difference equation orders k_1=1 and k_2=2, the matrices F,G,and H are

$$k\text{=1:}\ \ F = (I_m),\ \ G = (I_m),\ \ H_n = H_{1,n} = [y_{n-1}. \ . \ . \ ,y_{n-m}] \qquad (3.3.6)$$

$$k\text{=2:}\ \ F=\begin{bmatrix} 2I_m & -I_m \\ I_m & 0 \end{bmatrix},\ \ G = \begin{bmatrix} I_m \\ 0 \end{bmatrix},\ \ H_n = H_{2,n} = [H_{1,n}\ \ 0 \ . \ . \ . \ 0].$$

The input process noise w_n, the observation noise ϵ_n and the spectrum smoothness priors noise form the $2km+1$ vector $(w_n^T,\epsilon_n,\xi_n)^T$ that is assumed to be normally distributed and independent with time with the mean and covariance matrix,

$$\begin{pmatrix} w_n \\ \epsilon_n \\ \xi_n \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & S \end{pmatrix} \right) . \qquad (3.3.7)$$

In (3.3.7), the mxm diagonal matrix Q has the the element $1/\tau^2$ on the diagonal and for k_2=2,(in

(3.3.3)), the mxm diagonal matrix S has diagonal elements $1/(\lambda^2+\nu^2,...,m^4\lambda^2+\nu^2)$, (see (2.2.6)).

For a fixed difference order k_1, the best fit of the state space smoothness priors constraints-time varying AR coefficient model to the data $y_1, \ldots, y_N$, is the one for which the likelihood of the hyperparameters τ^2,λ^2,ν^2, are maximised. The likelihood is computed using the recursive formulas indicated in Section 3.2.

The Instantaneous Variance And The Instantaneous Spectrum

In many practical data analysis situations, the relatively fast wiggles of a nonstationary covariance time series appears to be modulated by a relatively slowly changing envelope function. The envelope function has an interpretation as a change of scale of the time varying AR coefficient model or equivalently as the smoothed (trend) value of the instantaneous variance, (Section 3.1). A key idea in that modeling is to find a variance stabilising transformation of the innovations that yields the instantaneous trend in an additive zero-mean constant variance observation noise, Wahba (1980). Let $s_n,n=1,...,N$ be a realisation of a zero-mean normally distributed white noise with unknown variance σ_n^2. Then, if $\sigma_{2n}^2=\sigma_{2n-1}^2$, $\chi_m^2=[s_{2m-1}^2+s_{2m}^2]/2$, is an independent sequence of chi-square random variables with two degrees of freedom. From Wahba (1980), the transformation $t_m=\log[\chi_m^2]+\gamma$, where $\gamma=0.57721$ is the Euler constant, leaves the independent random variable t_m with a distribution that is almost normal with the moments $E[t_m]=log\sigma_m^2$, $Var[t_m]=\pi^2/6$. That idea is exploited here.

Consider a second order difference equation constraint on the log variance defined by

$$\nabla^2 t_m = w_m. \tag{3.3.8}$$

In (3.3.8) $\{w_m\}$ is an independent zero-mean normally distributed sequence with unknown variance τ^2. Define a state vector by $z_m=[t_m\ t_{m-1}]^T$. Then in state space form the constraint model in (3.3.8) and the observation model are given by

$$x_m = \begin{bmatrix} -2 & 1 \\ 1 & 0 \end{bmatrix} x_{m-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} w_m \tag{3.3.9}$$

$$y_m = [1 \quad 0] x_m + \epsilon_m.$$

Application of the Kalman filter,prediction and smoothing algorithms described earlier yield the smooth value $t_{n|N}$, the logarithm of the smoothed estimate of the changing variance. Our estimate of the changing variance is $\sigma^2_{2m|N} = \sigma^2_{2m-1|N} = \exp(t_{m|N} + \gamma)$.

Motivated by earlier work on spectrum estimation, we define the instantaneous spectrum of a time varying coefficient AR process by

$$S_n(f) = \frac{\sigma_n^2}{\left| 1 - \sum_{k=1}^{m} a_{k,n} \exp(-2\pi i k f) \right|^2}; \quad -1/2 \leqslant f \leqslant 1/2. \tag{3.3.10}$$

The value of the instantaneous spectrum is obtained by substituting the smoothed estimates of the time varying AR coefficients and the smoothed estimate of the innovations variance σ_n^2 into (3.3.10).

3.4 EXAMPLES

A Nonstationary Mean Time Series Example

The RSWOMEN series of the Bureau of the Census data, (Zellner 1983), is analysed here. The series consists of the retail sales of women's apparel, reported in millions of dollars. The sales for each month are affected by the number of times each day of the week occurs, the trading-day effect, because buying behavior differs for each day of the week. (The sales are also affected by holidays.) We are interested in determining whether or not it is appropriate to include a trading day effect in the model and whether or not the globally stochastic AR component should be included in the model. The AIC statistic is used to determine the best of the alternative models.

The computations were done using the DECOMP.FORT program in TIMSAC-84 (Akaike et al. 1985). The model was fitted to the data $y_1, \ldots, y_{139}$, 24 data points were witheld. The AIC's for

AR model orders p=0,1,2,3, (as in (3.1.5)), respectively for the non trading day effect and trading day effect models are (111.51,96.96,98.80,98.65) and (88.07,68.34,67.23,68.91). An interpretation of those results is that models with an AR component, $p\neq 0$, are superior to models without AR components both with and without trading day effects and that the AIC best model,(AIC=67.23), is the trading day effect model with AR model order $p=2$. Figures 3a-e show selected computational responses for the non-AR component-trading day effect model. Figures 3f-j show selected computational results for the AR component-trading day effects model. The seasonal components, residual noise and trading day effects and seasonal plus trading effects are quite similar in appearance for both models.

Several aspects of the modeling results are noteworthy. The trend of the trend plus AR component model is smoother than the trend-non AR component model and the trend plus AR component is almost indistinguishable from the trend in the non AR component model. Also, the seasonal component is very regular whereas the seasonal plus trading day component reveals the expected slight irregularities.

A important property of the AIC best trend plus seasonal plus AR component model, instead of the trend plus seasonal model, can be seen in the (out-of-sample) forecasts for these models as shown in Figures 3e and 3j. In those illustrations we show the true series, the forecasted series and plus and minus one sigma of the forecast random variable. The plus or minus one sigma prediction intervals of the trend plus AR plus seasonal plus trading day components model is much tighter than the same quantity for the non AR component model. The increase in prediction variance per step in increasing horizon forecast, grows in accordance with the variance component terms in the matrix, in (3.1.8). The variance of the (wiggly) trend term in the non AR component model, is larger than the sum of variance terms of the (smooth) trend and AR component in the AR component model. That larger variance is reflected into the larger one sigma prediction interval of the non AR component model.

These results illustrate the flexibility of the decomposition of the nonstationary mean concept via smoothness priors modeling and the importance of the role of the AIC in selecting the best of alternative models.

Time Varying AR Coefficient Modeling, Nonstationary Covariance Time Series

The computations were realized using the TVCAR.FORT program in TIMSAC-84 (Akaike et al 1985). Figure 4a shows a seismic data event, $y_{1,...,}y_N$. The stochastic "background noise", P wave (the first abrupt change in the signal) and the S wave (the second abrupt change in the signal), are clearly discernable. Figures 4bd,f are graphs of computational results from the time varying AR coefficient model described in Section 3.3. Respectively they show the $log((y_{2m}^2 + y_{2m-1}^2)$, $m=1,...,N/2$ "unsmoothed envelope" data and the superimposed estimate of the envelope (changing variance), the evolution of the instantaneous power spectral density and the evolution of the partial correlation coefficients (parcors) of the fitted, AR order m = 8 time varying AR coefficient model. Figures 4c,e,g show the corresponding computational results from an "intervention analysis" model that is part of TVCAR.FORT.

The smoothed envelope for the non-intervention model is in fact quite smooth. Similarly the instantaneous spectrum and partial correlation coefficients (parcors) reflect the smooth transition in the time varying AR coeficients model, (3.3.2). Two visual inspection determined "outlier" events occur at $n=635$ and $n=1030$. They correspond to the arrival times of the P and S waves and are identified to the program, by human intervention, as large observation variance events. The large observation variance relaxes the priors constraint and permits the AR coefficients and subsequently derived quantities to change abruptly at those instants. The "validity" of this intervention type analysis is suggested by comparison of the results of the intervention type and non-intervention type analysis. The properties of the latter "drift" toward the former. The abrupt changes in the appearance of the envelope function and in the instantaneous spectra and parcors correspond to the physical interpretation that the P waves and S waves are different sources of energy at the observing seismometer.

3.5 COMMENTS, DISCUSSION

The paper by Akaike (1980) motivated our work in smoothness priors. In Akaike (1980), computations were done by a Householder transformation least squares algorithm of computational complexity $O(N^3)$. Brotherton and Gersch (1981) and Kitagwa (1981) demonstrated an equivalent state space modeling approach for the linear model with Gaussian system and observation noise. In the vicinity of the maximized likelihood, the likelihood is a rather flat function of the hyperparameters. This fact permits a relatively coarse grid discrete hyperparameter-likelihood search procedure to determine the values of the hyperparameters that tend to maximize the likelihood. Such a procedure preserves the $O(N)$ computational complexity inherent in the Kalman filter computations. A computational complexity of $O(N)$ version of that method was subsequently applied to a variety of nonstationary mean and nonstationary covariance time series modeling problems, (Gersch and Kitagawa 1983,1985, Kitagawa and Gersch 1984, 1985b). Variations of the procedures in those papers expressed in computer programs DECOMP.FORT and TVCAR.FORT (TIMSAC-84, Akaike et al. 1985), yielded the computational results shown here.

Potentially many more combinations and extensions of the models shown here for the modeling of nonstationary mean and nonstationary covariance time series by smoothness priors methods, are possible. For example, a generalization of the regression on trading days components, in the nonstationary mean-decomposition of time series modeling, could take into account constant coefficient and/or time varying coefficient regresssion on other time series. A time varying partial AR coefficients variation of the present time varying AR coefficient model for the modeling of nonstationary covariance time series, has already been implemented. Another potential variation on the time varying AR coefficient model would be to estimate the full nondiagonal mxm system noise covariance matrix (the matrix of hyperparameters). Gersch and Kitagawa (1985), an application of the time varying AR coefficient model, includes computation of the time varying covariance function. That computation is useful to permit computation of the mean square response to nonstationary excitation of building structures to single realisations of seismic event data.

Some other linear model-Gaussian distrurbances- state space smoothness priors models have been implemented. Kitagawa and Takanami (1985) show a smoothness priors modeling method for the extraction of seismic signals from correlated background noise. The smoothness priors innovation in that work is the implementation of a non constant or time varying hyperparameter. That hyperparameter achieves a time varying balance of the tradeoff between the variances of the seismic signal and the background noise. The modeling of continuous model time series with discrete time observations is another domain where smoothness priors state space modeling has been exhibited. Kitagawa (1984) includes a smoothness priors variation of the Jones (1980) continuous time AR process-discrete time observations modeling. The application in Kitagawa (1984) is to irregularly spaced or missing data time series modeling.

4. STATE SPACE NON GAUSSIAN MODELING OF NONSTATIONARY MEAN TIME SERIES

A non-Gaussian state space approach to the modeling of nonstationary time series is shown. Neither the system noise nor the observation noise need be Gaussian. Recursive formulas for the prediction, filtering and smoothing of the state are given. A numerical method, based on a piecewise linear approximation to the density functions for realising these formulas, is also given. The merits and potential wide applicability of this approach to non-Gaussian modeling are illustrated by some numerical examples. Extension of this method to the state space modeling of nonlinear systems is straightforward.

Earlier in this chapter we demonstrated the wide range of applicability of the linear model with Gaussian system and observation noise. There are numerous problems for which Gaussian modeling is inadequate. For example, the problem of trend estimation becomes difficult when the trend has discontinuities as well as smooth changes and when there are observation outliers. A simple linear Gaussian model with small process noise variance does not track jumps or discontinuities very well. A model with large process noise variance will respond to sudden changes in the trend but it will also be inappropriately wiggley where the trend is quite smooth. The treatmeant of such trend discontinuities with the included possibility of observation outliers in the linear Gaussian model framework requires a complicated model. Heavy tailed distributions for process and observation noise can cope with these problems with a simple model. Also, smoothing problems in which there is a time varying variance and/or a nonhomogenous binomial or Poisson mean require a non Gaussian system noise model formulation. Similarly nonlinear models such as storage models for riverflow and a ship's nonlinear manueverability require non Gaussian distribution models.

Thus, the development of methods for treating systems with non Gaussian distributions is well motivated. In earlier attempts, systems with non Gaussian distributions were approximated by the use of extended Kalman filters, sums of Gaussian distributions , by Edgeworth or Gram-

Charlier expansions etc.. (See Alspach and Sorenson 1972, for example.) Here, we approximate the probability density functions directly by a piecewise linear function. The recursive prediction, filtering and smoothing computation required by the state space modeling are realized by numerical integration. A similar approach was considered by Bucy and Senne (1971) and de Figueiredo and Jan (1971) in the context of nonlinear filtering problems. Such an approach is more feasible now than it was several years ago because of the development and proliferation of fast computational facilities. In Section 4.1 the state space prediction , filtering and smoothing formula aspects of the numerical computations are derived and the computation of the likelihood for the not necessarily Gaussian distribution model are shown. Numerical examples are shown in Section 4.2 and a discussion and comments are in Section 4.3.

4.1 THE NON GAUSSIAN STATE SPACE MODEL

Consider the stationary state space system described by

$$\begin{aligned} x_n &= Fx_{n-1} + Gw_{n-1} \\ y_n &= Hx_n + \epsilon_n \end{aligned} \tag{4.1.1}$$

where as before F, G, and H are linear transformations. The independent and independent of each other, but not necesssarily Gaussian process and observation noises are w_n and ϵ_n respectively. The initial state vector x_0 is distributed in accordance with $p(x_0)$ and the conditional density of the state at time n, given the obervations $(y_1,...,y_m) = Y_m$ is denoted by $p(x_n| Y_m)$. Then, the recursive formulas for the one step ahead prediction,filtering and smoothing densities are derived as follows:

One step ahead prediction (time update)

$$\begin{aligned} p(x_n| Y_{n-1}) &= \int_{-\infty}^{\infty} p(x_n,x_{n-1}| Y_{n-1})dx_{n-1} \\ &= \int_{-\infty}^{\infty} p(x_n| x_{n-1})p(x_{n-1}| Y_{n-1})dx_{n-1} \end{aligned} \tag{4.1.2}$$

Filtering (observation update)

$$\begin{aligned} p(x_n|\,Y_n) = p(x_n|\,y_n, Y_{n-1}) &= p(x_n, y_n|\,Y_{n-1})/p(y_n|\,Y_{n-1}) \qquad (4.1.3) \\ &= p(y_n|\,x_n)p(x_n|\,Y_{n-1})/p(y_n|\,Y_{n-1}) \end{aligned}$$

where $p(x_n|\,x_{n-1})$ is the density of x_n given the previous state vector x_{n-1}, $p(y_n|\,x_n)$ is the density of y_n given x_n and $p(y_n|\,Y_{n-1})$ is obtained by $\int p(y_n|\,x_n)p(x_n|\,Y_{n-1})dx_n$.

Similarly, consider the expression for the joint density of x_n and x_{n+1}, given the entire observation sequence Y_N,

$$\begin{aligned} p(x_n, x_{n+1}|\,Y_N) &= p(x_{n+1}|\,Y_N)p(x_n|\,x_{n+1}, Y_n) \qquad (4.1.4) \\ &= p(x_{n+1}|\,Y_N)p(x_n, x_{n+1}|\,Y_n)/p(x_{n+1}|\,Y_n) \\ &= p(x_{n+1}|\,Y_N)p(x_{n+1}|\,x_n)p(x_n|\,Y_n)/p(x_{n+1}|\,Y_n) \end{aligned}$$

From (4.1.4) we obtain the formula for smoothing:

$$\begin{aligned} p(x_n|\,Y_N) &= \int_{-\infty}^{\infty} p(x_n, x_{n+1}|\,Y_N)dx_{n+1} \qquad (4.1.5) \\ &= p(x_n|\,Y_n)\int_{-\infty}^{\infty} p(x_{n+1}|\,Y_N)p(x_{n+1}|\,x_n)/p(x_{n+1}|\,Y_n)dx_{n+1}. \end{aligned}$$

In the linear Gaussian case, the conditional densities $p(x_n|\,Y_{n-1})$, $p(x_n|\,Y_N)$ and $p(x_n|\,Y_N)$ are characterized by mean vectors and covariance matrices . Correspondingly, (4.1.2),(4.1.3) and (4.1.5) lead to the Kalman filter and the fixed interval smoothing algorithm. In the non Gaussian or nonlinear case however, it is necessary to evaluate the non Gaussian densities explicitly at each step. The algorithms above, (4.1.3)-(4.1.5) can be realised numerically by piecewise linear approximations to the density functions, transformation of densities, convolution of densities and Bayes theorem (product of two densities and normalisation). Details of the numerical computations are in Kitagawa (1987).

In general,the non Gaussian model has some unknown parameters. The best choice of the parameters can be found by maximizing the log likelihood defined by

$$l(\theta) = \log p(y_1,\ldots,y_N) = \sum_{n=1}^{N} \log p(y_n | y_1,\ldots,y_{n-1}) = \sum_{n=1}^{N} \log p(y_n | Y_{n-1}). \qquad (4.1.6)$$

The term $p(y_n | Y_{n-1})$ appears in (4.1.3) and can be evaluated numerically. If we have several candidate models, including models with different types of system noise or observation noise density functions, we choose the model for which the AIC is minimum.

4.2 NUMERICAL EXAMPLES

Estimation of a Shifting Mean Value

Consider the data simulated from the following model,

$$Y_n \sim N(\mu_n,1)$$

$$\mu_n = \begin{cases} 0 & n= \ \ \ 1,\ldots,100 \\ 1 & n=101,\ldots,200 \\ -1 & n=201,\ldots,300 \\ 2 & n=301,\ldots,400 \end{cases} \qquad (4.2.1)$$

The data is shown in Figure 5a. The problem is to estimate the abruptly changing mean value function μ_n.

For this type of data we used the model

$$\nabla^k t_n = w_n \qquad (4.2.2)$$

$$y_n = t_n + \epsilon_n.$$

As before, ∇ is the difference operator defined by $\nabla t_n = t_n - t_{n-1}$ and w_n and ϵ_n are white noise sequences that are not necessarily normally distributed. For simplicity we assume that the difference order k is one. Equation (4.2.2) is a special form of the state space model, Section (3.1.), with $x_n = t_n, F = G = H = 1$. We considered the following model classes:

$$Model(a): \quad w_n \sim aN(0,\tau^2)+(1-a)N(0,\tau_s^2), \quad \epsilon_n \sim N(0,1) \qquad (4.2.3)$$

$$Model(b): \quad w_n \sim Q(b,\tau^2), \quad \epsilon_n \sim N(0,1)$$

Model(a) denotes a mixture of Gaussian system noises. In (4.2.3), for *Model*(a) and *Model*(b),

$N(0,r)$ denotes the Gaussian distribution with mean 0 and variance r. In *Model*(b), $Q(b,\tau^2)$ denotes the distribution of the Pearson system with density $q(x;b,\tau^2)=C(\tau^2+x^2)^{-b}$ with $\frac{1}{2}<b\leqslant\infty$ and $C=\tau^{2b-1}\Gamma(b)/\Gamma(\frac{1}{2})$. This family links the Cauchy distribution $(b=1)$ and the Gaussian distribution $(b=\infty)$. In *Model*(a), τ_s^2 was arbitrarily set to 4.0, approximately the sample variance of the simulated data. The maximum likelihood estimate of τ^2 for the Gaussian model, *Model*(a), with $a=1.0$ or equivalently *Model*(b) with $b=\infty$, was $\hat{\tau}^2$=0.0429. The AIC of the model was 1240.33. For the mixture of Gaussian system noises model, $\hat{a}$=0.989, $\hat{\tau}^2$=0.0000014 and AIC=1212.48. We tried four Pearson family models: b=0.6,0.8,1.0 and ∞. b=0.80 is the AIC best Pearson family model with $\hat{\tau}^2$=0.000002 and AIC=1215.20. The AIC best model is the mixture of Gaussian system noises model.

Figure 5b-5d shows the marginal posterior density $p(x_n|Y_N)$ versus time n for the Gaussian model, the best Pearson system model and the mixture of Gaussians model respectively. For the Gaussian model, Figure 5b, the densities obtained have identical shape except for the ends of the time interval where the densities become slightly broader. In Figure 5c, the shape of the posterior density varies with time. When the mean value shifts, the density becomes heavy tailed on one side. The Gaussian mixtures model also exhibits the latter behavior.

Figures 5e-5g shows the mean (bold) and ± 1,2,3 sigma intervals of the $p(x_n|Y_N)$ versus n for the Gaussian model and the median (bold) and corresponding 0.13,2.27,15.87,84.13,97.73 and 99.87 percentage points of the Pearson system model and the Gaussian mixtures model respectively. For the Gaussian model, the estimated mean value function becomes a wiggly curve and does not reflect the sudden change of the mean value. The estimated median of the Pearson system and the Gaussian mixtures models do capture the sudden change of the signal mean value. The multimodal or skewed distribution and jumps of the mean value are typical of the phenomenon that are seen in non Gaussian modeling.

Estimation Of Changing Variance

The estimation of changing variance was discussed in Section 3.3.3 in the context of fitting a time varying AR coefficient model to a seismic signal. The same idea is exploited here to estimate the changing variance of the same seismic signal with the state space non Gaussian modeling method. A first order difference model for the trend of the log of the sum of the squares of successive observations was used, $\nabla t_m = t_m - t_{m-1} \sim q(b, \tau^2)$, $y_m = t_m + \epsilon_m$ with $m=1,..,N/2$. Two models were considered, the Gaussian system noise-Gaussian observation noise and the Cauchy system noise with an $r(x) = exp(x - exp^x)$ observation noise model. The corresponding AIC's were 4778.94 and 4222.84. The latter model was the AIC best model. The original seismic wave $y_n, n=1,...,N$ and the $log((y_{2m}^2 + y_{2m-1}^2)/2)$, $m=1,...,N/2$ signals are shown in Figures 6a and 6b respectively. The Gaussian model smoothed mean and $\pm 1,2,3\sigma$ and non Gaussian model smoothed median and corresponding probability point curves are shown in Figures 6c,6d respectively. Those illustrations indicate that the Cauchy system noise model yields better estimates of the smooth mean and abrupt changes of the mean innovations variance than the simple Gaussian system noise model. Modeling the real data changing variance seismic signal with the state space non Gaussian system noise method automatically yields the location of abrupt changes in the mean of the siganal.

4.3 COMMENTS, DISCUSSION

The results shown here were obtained using a simple one dimensional state vector. In principle, it is straightforward to extend the computational formulas to higher dimensional state systems. The resulting increase in the computational burden required to compute the convolution of density functions becomes quite severe. A variety of numerical techniques have been investigated to cope with this problem. Very likely the use of more powerful computers rather than the increase of effort in numerical analysis methods will be more expeditious in the development of the non Gaussian state space smoothness priors method.

Several other problems lend themselves to the application of the one dimensional non Gaussian modeling shown here. Kitagawa (1987) shows an application to the handling of discrete distributions. The time varying mean of a real, nonstationary (nonhomogeneous) binary process, is estimated. Also, smoothing of the log periodogram using a state space model with $w_n \sim \log\chi^2$ and ϵ_n~Cauchy or ϵ_n~ Gaussian is an alternative to the Gaussian distributions approach in Wahba (1980).

Our procedure also extends quite naturally to the analysis of nonlinear systems. The one-step-ahead prediction formula (4.1.2) and the filtering formula (4.1.3) are applicable even for nonlinear systems.

Time series with nonstationarities, nonlinearities and outliers that have been difficult to analyze by conventional linear Gaussian models can be quite simply analyzed with the non Gaussian model. The computational burden using the non Gaussian filter and smoother is substantial. The develpment of faster algorith.ns and the use of faster computers will alleviate this burden.

5. SUMMARY

The ingredients for the smoothness priors analysis of time series are the model, the prescription of the priors, the criterion for goodness of model fit and the computational method.

Initially, smoothness priors modeling of stationary, nonstationary mean, and nonstationary covariance time series was demonstrated in the context of the linear model with Gaussian distributed system noise and with Gaussian distributed observation noise. Both time domain and frequency domain specifications of the prior distribution of the model parameters were considered. A hyperparameter specifies the degree of belief in the prior distribution. The smoothness priors method of analysis derives its unity from the fact that the likelihood of the Bayesian model (the likelihood of the hyperparameter(s)) is the single criterion by which the goodness of fit of the model is determined. The maximization of the likelihood of a small number of hyperparameters permits the modeling of time series with complex structure and a large number of implicitly inferrred parameters. When there are alternative candidate smoothness priors models, we use Akaike's AIC to determine the best of alternative models, (the likelihood of the model has a central role in the AIC). Householder transformation least squares and Kalman filter algorithms, were the means for the realization of the smoothness priors time series modeling.

Finally, we demonstrated a state space representation not necessarily Gaussian not necessarily linear model method of smoothness priors modeling. In that method, piecewise-linear approximation to densities and numerical integration computations were employed. Conceptually all of the possible combinations of models and smoothness priors computations could be realized with this method.

The extensive applicability of smoothness priors modeling methods in time series modeling was demonstrated. A large number of other problems that have not been well solved by more traditional time series methods remain to be solved by that method.

REFERENCES

Akaike, H. (1973), Information Theory and an Extension of the Maximum Likelihood Principle, in Second International Symposium in Information Theory, B.N. Petroc and F. Caski eds. Budapest, Akademiai Kiado, 267-281.

Akaike, H.(1979), Smoothness Priors and the Distributed Lag Estimator, Dept. Statistics,Stanford Univ., Stanford, CA., T.A. Anderson, Project Director, Tech. Report 40.

Akaike, H. (1980), Likelihood and the Bayes procedure, In, Bayesian Statistics, J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith eds., University Press, Valencia , Spain, 143-166.

Akaike, H. (1980), Seasonal Adjustment by a Bayesian Modeling, Journal of Time Series Analysis, 1,1-13.

Akaike,H., Ozaki,T.,Ishiguro,M.,Ogata, Y.,Kitagawa, G.,Tamura, Y-H.,Arahata, E., Katsura, K.,and Tamura,Y.,(1985), TIMSAC-84,Part 1 and Part 2, Vols. 22 and 23, Computer Science Monographs, The Institute of Statistical Mathematics, Tokyo,Japan.

Alspach, D.L. and Sorenson, H.W.(1972), Nonlinear Bayesian Estimation using Gaussian Sum Approximations, IEEE Transactions on Automatic Control, AC-17,439-448.

Anderson, B.D.O. and Moore, J.B. (1979), Optimal Filtering, Prentice Hall, New York.

Askar, M. and Derin, H. (1981), A Recursive Algorithm for the Bayes Solution of the Smoothing Problem, IEEE Transactions on Automatic Control, AC-26,558-561.

Astrom K.J.A and Whittenmark B. (1971), Problems of Identification and Control, Journal of Mathematical Anaysis and Applications,34,90-113.

Berger J. O. (1985), Statistical Decision Theory, Foundations, Concepts and Methods, 2nd edition, New York, Springer-Verlag

Bohlin T. (1976), Four Cases of Identification of Changing Systems, in System Identiifaction Advances and Case Studies, R. Mehra and D.G. Lainotis,eds., New York, Academic Press.

Box G.E.P and Jenkins G.M.(1970),Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco.

Brotherton T. and Gersch W. (1981), A Data Analytic Approach to the Smoothing Problem and Some of its Variations, Proceedings of the 20th IEEE Conference on Decision and Control, 1061-1069.

Bucy, R.S. and Senne K.D.(1971), Digital synthesis of nonlinear filters, Automatica,7,287-289.

Campbell, M.J. and Walker, A.M. (1977), A Survey of Statistical Work on the McKensie River Series of Annual Canadian Lynx Trappings for the Years 1821-1934 and a New Analysis, Journal of the Royal Statistical Society, A 140, 411-431.

de Figueiredo, J.R.P and Jan, Y.G. (1971), Spline filters, Proceedings of the 2nd Symposium on Nonlinear Estimation Theory and its Applications, San Diego, 127-141.

Gersch, W. and Kitagawa, G. (1983), The Prediction of Time Series with Trends and

Seasonalities, Journal of Business and Economic Statistics,1,253-264.

Gersch, W. and Kitagawa, G. (1984), A Smoothness Priors Method for Transfer Function Estimation, Proceedings of the 23rd IEEE Conference On Decision and Control, 363-367.

Gersch W. and Kitagawa G (1985), A Time Varying AR Coefficient Model for Modeling and Simulating Earthquake Ground Motion, Earthquake Engineering and Structural Dynamics,13,243-254.

Good I.J. (1965), The Estimation of Probabilities, Cambridge, Mass., MIT Press.

Good I.J. and Gaskins J.R. (1980), Density Estimation and Bump Hunting by the Penalised Likelihood Method Exemplified by Scattering and Meteorite Data,75,42-73 Journal of the Ameriacn Statistical Association,75,42-73.

Ishiguro M. and Sakamoto Y. (1983), A Bayesian Approach To Binary Respone Curve Estimation, Annals of the Institute of Statistical Mathematics,35,B,115-137.

Ishiguo M, Akaike H., Doe M. and Nakai S.(1981) A Bayesian Approach to the Analysis of Earth Tides, Proc. 9th Inst. Conference on Earth Tides.

Jones R.H. (1980), Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations, Technometrics,22,389-395.

Kalman R.E.K. (1960), A New Approach to Linear Filtering and Prediction Problems, Transactions of ASME, Journal of Basic Engineering, 82D,35-45.

Kesler, S.B. (1986), Modern Spectrum Analysis II, (editor), IEEE Press, New York.

Kitagawa, G. (1981), A Nonstationary Time Series Model and its Fitting by a Recursive Filter, Journal of Time Series Analysis,2,103-116.

Kitagawa, G. (1983), Changing Spectrum Estimation, Journal of Sound and Vibration, 89, No. 4, 443-445.

Kitagawa G. (1984), State Space Modeling of Nonstationary Time Series and Smoothing of Unequally Spaced Data, in Time Series Analysis of Irregularly Observed data, E.Parzen, ed.,New York, Springer-Verlag, Lecture Notes in Statistics, 25,189-210.

Kitagawa, G. (1987), Non-Gaussian State Space Modeling of Nonstationary Time Series,(with discussion), Journal of the American Statistical Association, 82, to appear.

Kitagawa G.and Gersch W. (1984), A Smoothness Priors-State Space Modeling of Time Series with Trend and Seasonality, Journal of the American Statistical Association, 79, 378-389.

Kitagawa G.and Gersch W. (1985a), A Smoothness Priors Long AR Model Method for Spectral Estimation, IEEE Transactions on Automatic Control, AC-30,57-65.

Kitagawa G.and Gersch W. (1985b), A Smoothness Priors Time Varying AR Coefficient Modeling of Nonstationary Time Series, IEEE Transactions on Automatic Control, AC-30,48-56.

Kitagawa G.and Takanami T. (1985), Extraction of Signal by a Time Series Model and Screening Out Micro Earthquakes, Signal Processing,8,303-314.

Kohn R.and Ansley C.R. (1987), Smoothness Priors and Optimal Interpolation and Smoothing, in Bayesian Analysis of Time Series and Dynamic Systems, J.C. Spall,ed., Marcel Dekker, New York.

Kozin F. (1977), Estimation and Modeling of Nonstationary Time Series, in Proceedings Syposium in Applied Computaional Methods in Engineering, Univ. Southern California, Los Angeles,603-612.

Lindley D.V. and Smith A.F.M. (1972), Bayes Estimate for the Linear Model, Journal of the Royal Statistical Society,B,34,1-41.

Nakamura T.(1986), Bayesian Cohort Models for General Cohort Table Analysis, Annals of the Institute of Statistical Mathematics, 38, Part B, 353-370.

Naniwa, S. (1986), Trend Estimation via Smoothness Priors-State Space Modeling, Monetary and Economic Studies, Institute for Monetary and Economic Studies, Bank of Japan, 4,79-112.

Nicholls and Quinn (1985) Time Series in the Time Domain, in Handbook of Staistics, eds. E.J. Hannan and P.R. Krishnaiah, Amsterdam, Horth Holland.

O'Sullivan F., Yandell B.S. and Raynor W.J. Jr. (1986), Automatic Smoothing of Regression Functions in Generalized Linear Models, Journal of the American Statistical Association,81,96-103.

Shiller R. (1973), A Distributed Lag Estimator Derived from Smoothness Priors, Econometrica, 41,775-778.

Tanabe K. and Sagae (1987) Smoothness priors density estimation, in preparation.

Tikhonov A. N. (1965), Incorrect Problems of Linear Algebra and a Stable Method for their Solution. Soviet Math. Dokl. 6, 988-991.

Titterington D.M. (1985), Common Structure of Smoothing Techniques in Statistics, International Statistical Review,53,141-170.

Vinod H.D. and Ullah J.L. (1981), Recent Advances in Regression Methods, New York, Marcel Dekker.

Wahba G. (1977) A Survey of Some Smoothing Problems and the Method of Generalized Cross-Validation for Solving Them, in: Applications of Statistics, ed P.R. Krishnaih, North Holland, 507-524.

Wahba G. (1980), Automatic Smoothing of the Log Periodogram, Journal of the American Statistical Association,75,122-132.

Wahba G. (1982),Constrained Regularization for Ill Posed Linear Operator Equations with Applications in Meteorology and Medicine, Statistical Theory and Related Topics III,2,eds S.S. Gupta and J.O. Berger, New York, Academic Press,383-418.

Wahba G. (1983) Bayesian confidence intervals for the cross-validated smoothing spline, Journal of the Royal Statistical Society B,45,133-150.

Wecker W.E. and Ansley C.R. (1983), The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing, Journal of the American Statistical Association,78,81-89.

Whittaker E.T. (1923), On a New Method of Graduation, Proceedings of the Edinborough Mathematical Association,78,81-89.

Whittaker E.T. and Robinson G. (1924), Calculus of Observations, A Treasure on Numerical Calculations, Balackie and Son, Lmtd., London. pp 303-306

Whittle P. (1965), Recursive Relations for Predictors of Non-stationary Processes, Journal of the Royal Statistical Society,27,B,523-532.

Zellner A. (1983), Editor, Applied Time Series Analysis of Economic Data, Washington D.C., U.S. Department of Commerce, Bureau of the Census.

LEGENDS

FIGURE 1. Trend Estimation

a: Truncated Gaussian signal and signal plus noise, b: Signal plus noise plus smoothed trend with a too large hyperparameter, c: Signal plus noise plus smoothed trend with a too small hyperparameter, d: Signal plus noise plus smoothed trend with optimum hyperparameter,

Figure 2. (2.1)Spectral Densities From Canadian Lynx Data Example

a: Spectral density versus frequency, smoothness priors model, b: Spectral density versus frequency, AIC-AR model, c: Superposition of spectral densities versus frequency, AR models.

FIGURE 3. (3.1)Nonstationary Mean, RSWOMEN Data.

Trend plus seasonal plus trading component model, a,b,c,d,e.

a: Original data and trend, b: Seasonal component, c: Trading day effect, d: Seasonal plus trading day effect, e: True, predicited and plus and minus one sigma plus predicted.

Trend plus seasonal plus AR plus trading component model, f,g,h,i,j. f: Original data and trend, g: AR component, h: Original plus trend plus AR component, i: Residual noise, j: True, predicited and plus and minus one sigma plus predicted.

FIGURE 4. Nonstationary Covariance, Seismic Data.

a. Seismic data,$y_1,...,y_N$. "ordinary model" b,d,f; "intervention model" c,e,g.

b,c: $log((y^2_{2m} + y^2_{2m-1})/2)$, $m=1,...,N/2$ data and smoothed envelope, d,e: instantaneous power spectral density, f,g: parcors.

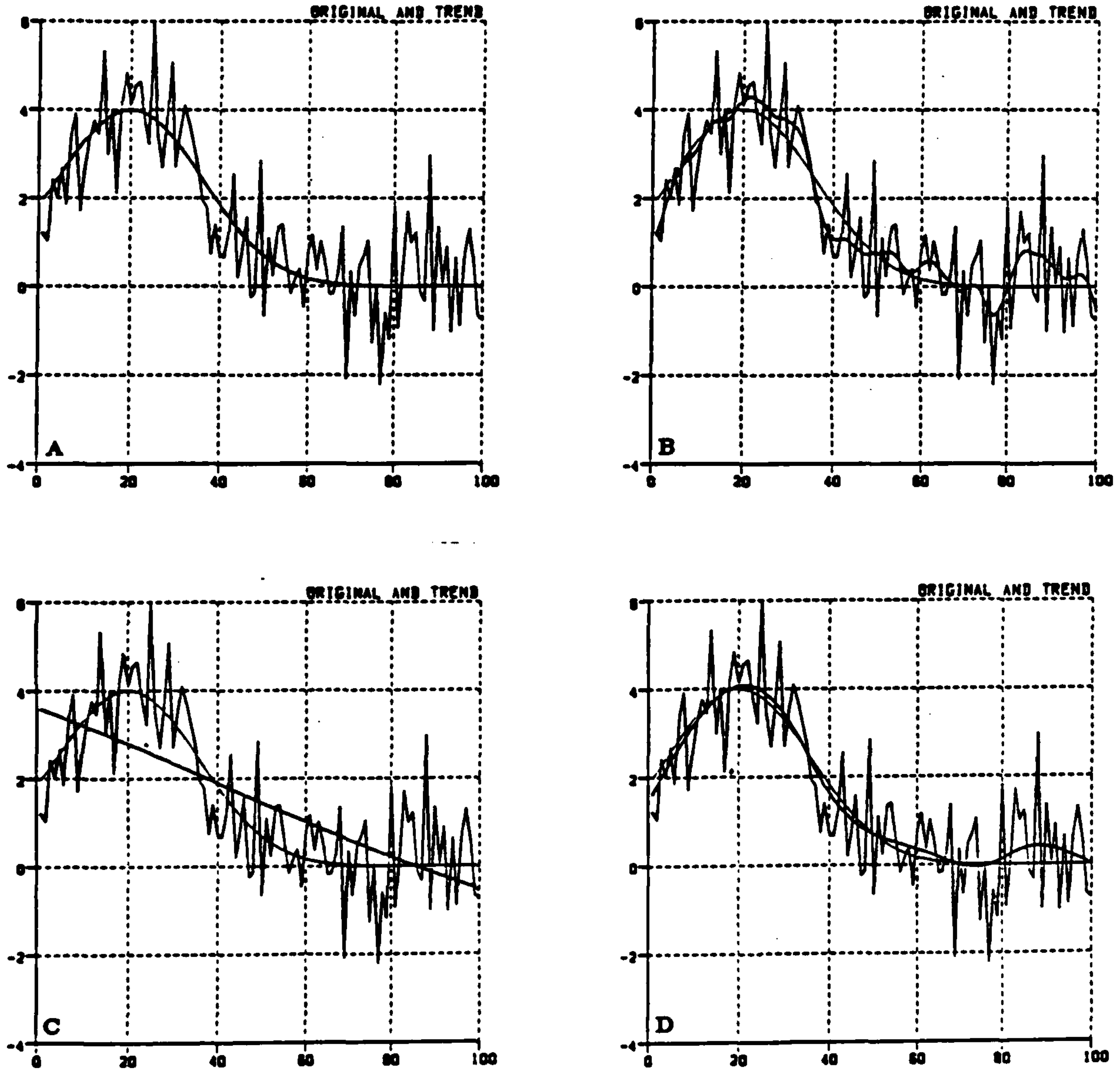
FIGURE 5. State Space Model Non Gaussian Discontinuous Trend Example.

a: Abruptly changing trend data, b: Smoothed state estimate, Gaussian system noise model, c: Smoothed state estimate, Pearson system-system noise model, d: Smoothed state estimate, Gaussian mixture system noise smodel, e: Posterior mean, $\pm 1,2,3\sigma$, Gaussian system noise model, f: Posterior median and (0.13,2.27,15.87,84.13,97.73) percentage points, Pearson system-system noise model. g: Posterior median and (0.13,2.27,15.87,84.13,97.73) percentage points, Gaussian system noise model.

FIGURE 6. State Space Model Non Gaussian Envelope of Seismic Signal Example.

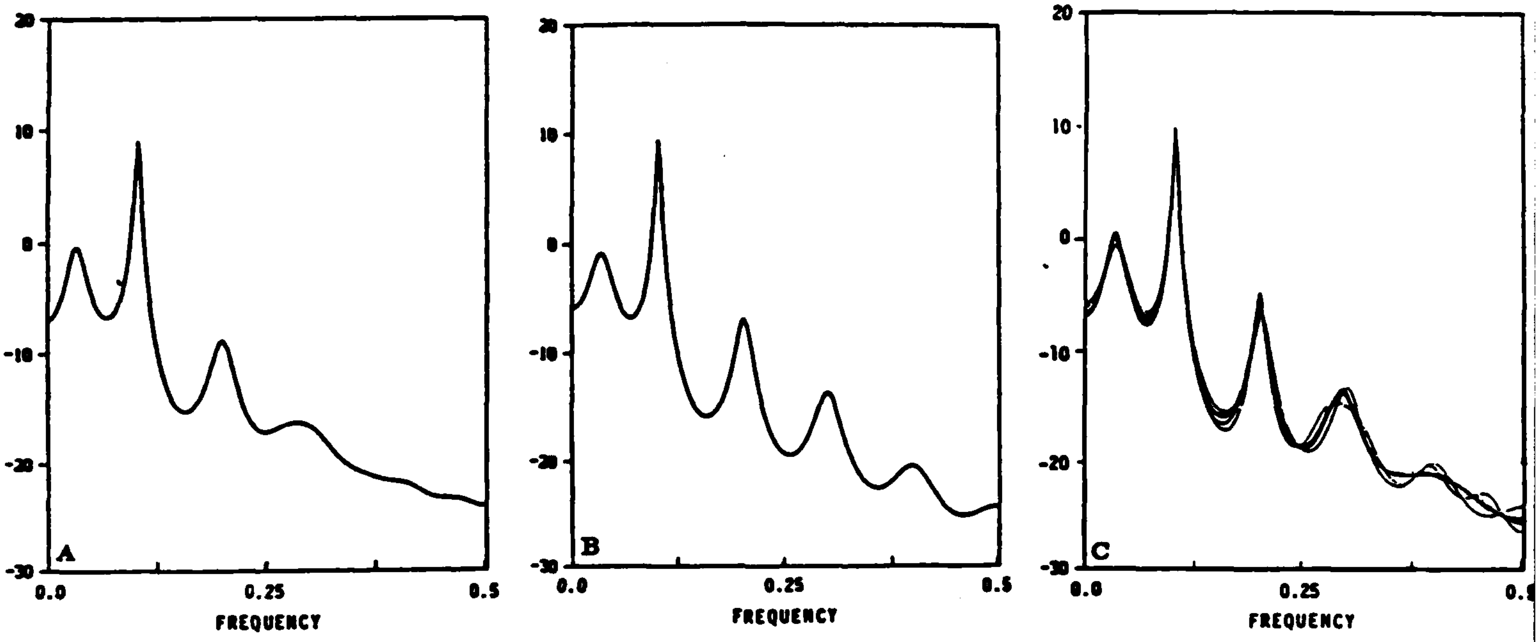
a: Seismic data,$y_1,\ldots,y_N$. b: $log((y_{2m}^2 + y_{2m-1}^2)/2)$, $m=1,\ldots,N/2$ "envelope" data, c: Posterior mean, $\pm 1,2,3\sigma$, Gaussian disturbances model, d: Posterior mode, (0.13,2.27,15.87,84.13,97.73) percentage points, non-Gaussian disturbances model.

FIGURE



ORIGINAL AND TREND
6
4
2
0
-2
-4
0
20
40
60
80
100
A
ORIGINAL AND TREND
6
4
2
0
-2
-4
0
20
40
60
80
100
B
ORIGINAL AND TREND
6
4
2
0
-2
-4
0
20
40
60
80
100
C
ORIGINAL AND TREND
6
4
2
0
-2
-4
0
20
40
60
80
100
D

FIGURE 2



20
10
0
-10
-20
-30
A
0.0
0.25
0.5
FREQUENCY
20
10
0
-10
-20
-30
B
0.0
0.25
0.5
FREQUENCY
20
10
0
-10
-20
-30
C
0.0
0.25
0.
FREQUENCY

FIGURE 3

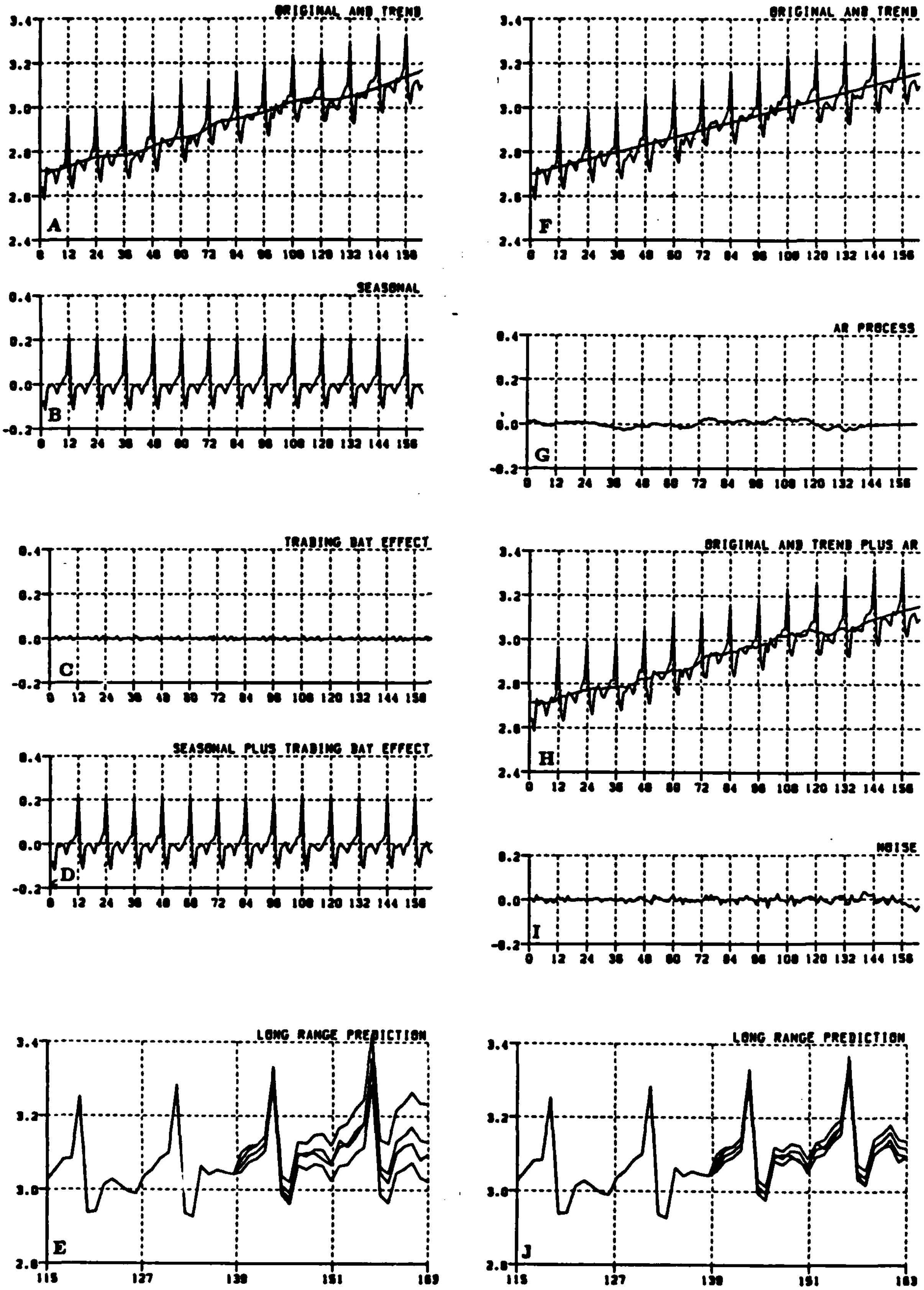

ORIGINAL AND TREND
3.4
3.2
3.0
2.8
2.6
2.4
A
0
12
24
36
48
60
72
84
96
108
120
132
144
156
SEASONAL
0.4
0.2
0.0
-0.2
B
TRADING DAY EFFECT
C
SEASONAL PLUS TRADING DAY EFFECT
D
LONG RANGE PREDICTION
E
2.8
115
127
139
151
163
ORIGINAL AND TREND
F
AR PROCESS
G
ORIGINAL AND TREND PLUS AR
H
NOISE
I
LONG RANGE PREDICTION
J

FIGURE 4

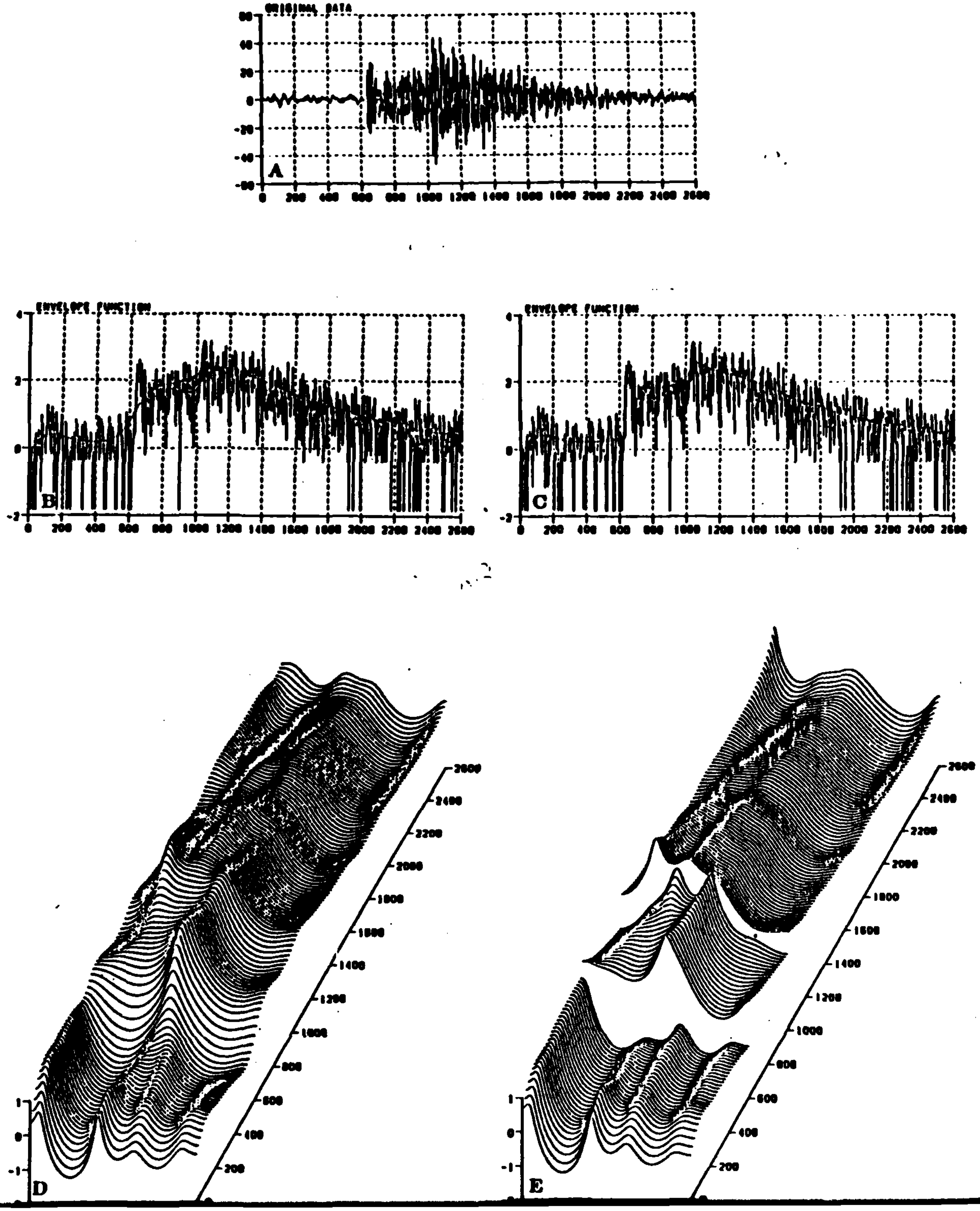

ORIGINAL DATA
60
40
20
0
-20
-40
-60
A
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
ENVELOPE FUNCTION
4
2
0
-2
B
ENVELOPE FUNCTION
C
1
0
-1
D
E
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600

FIGURE 4 CONTINUED

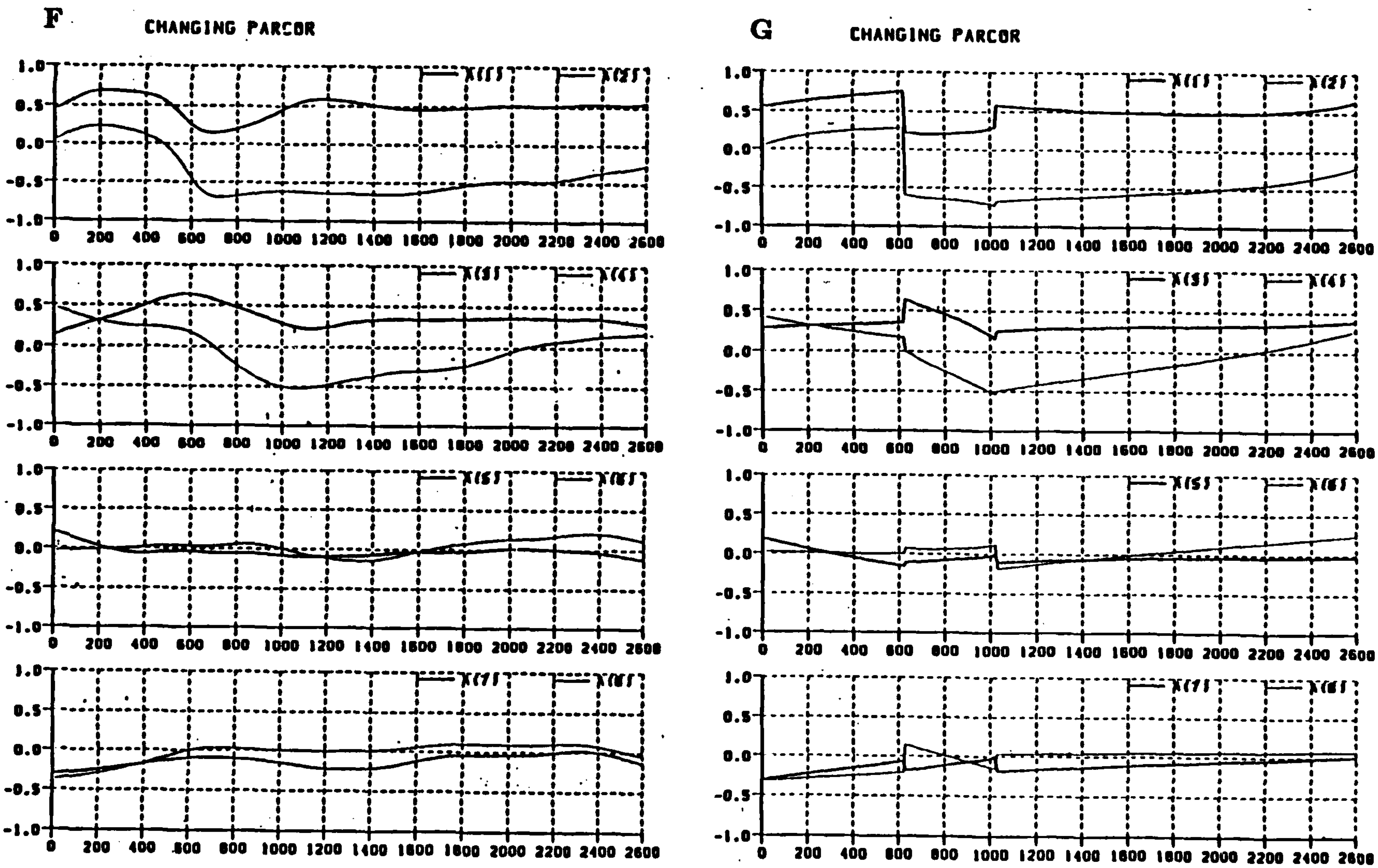


FIGURE 5

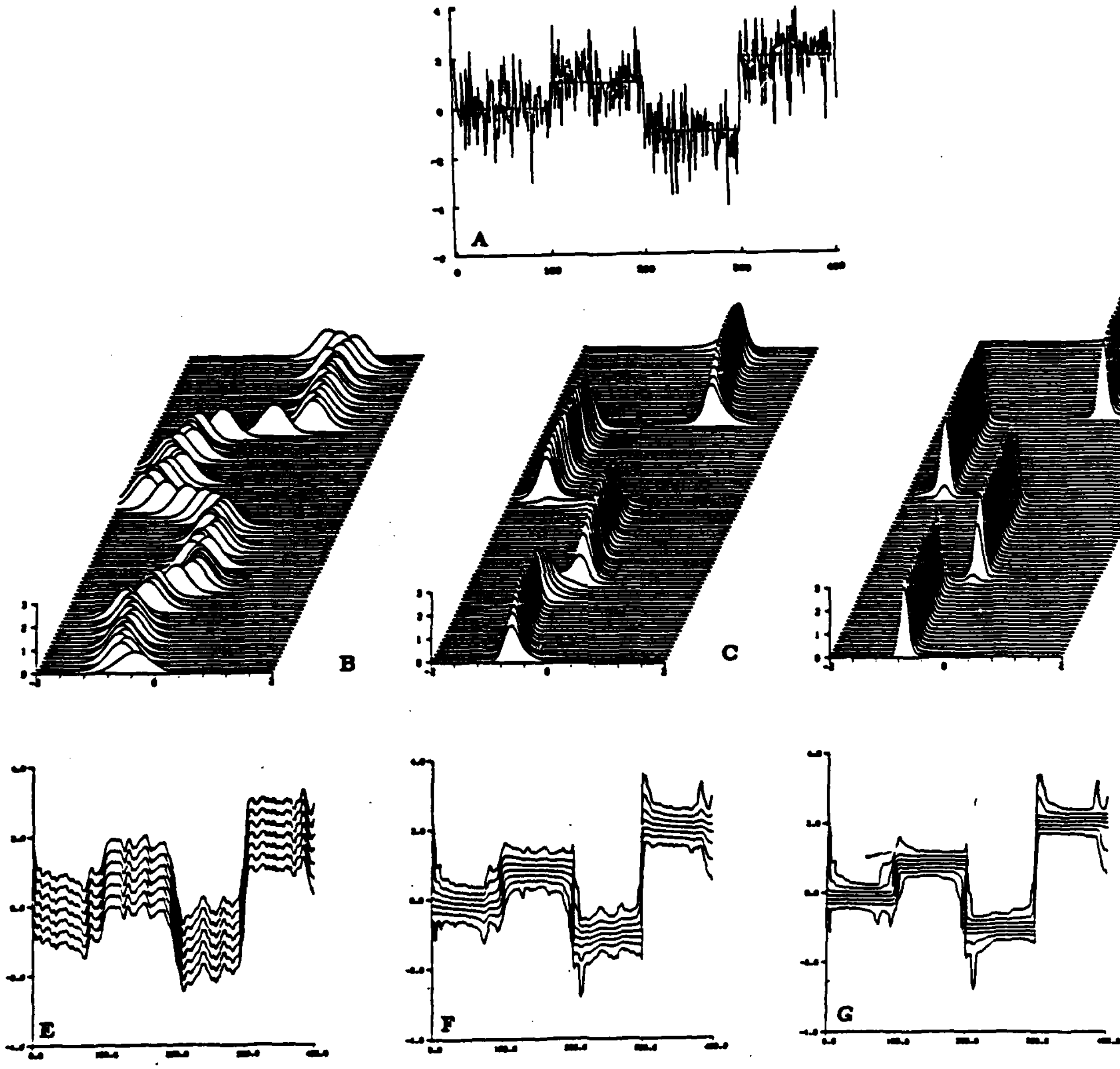
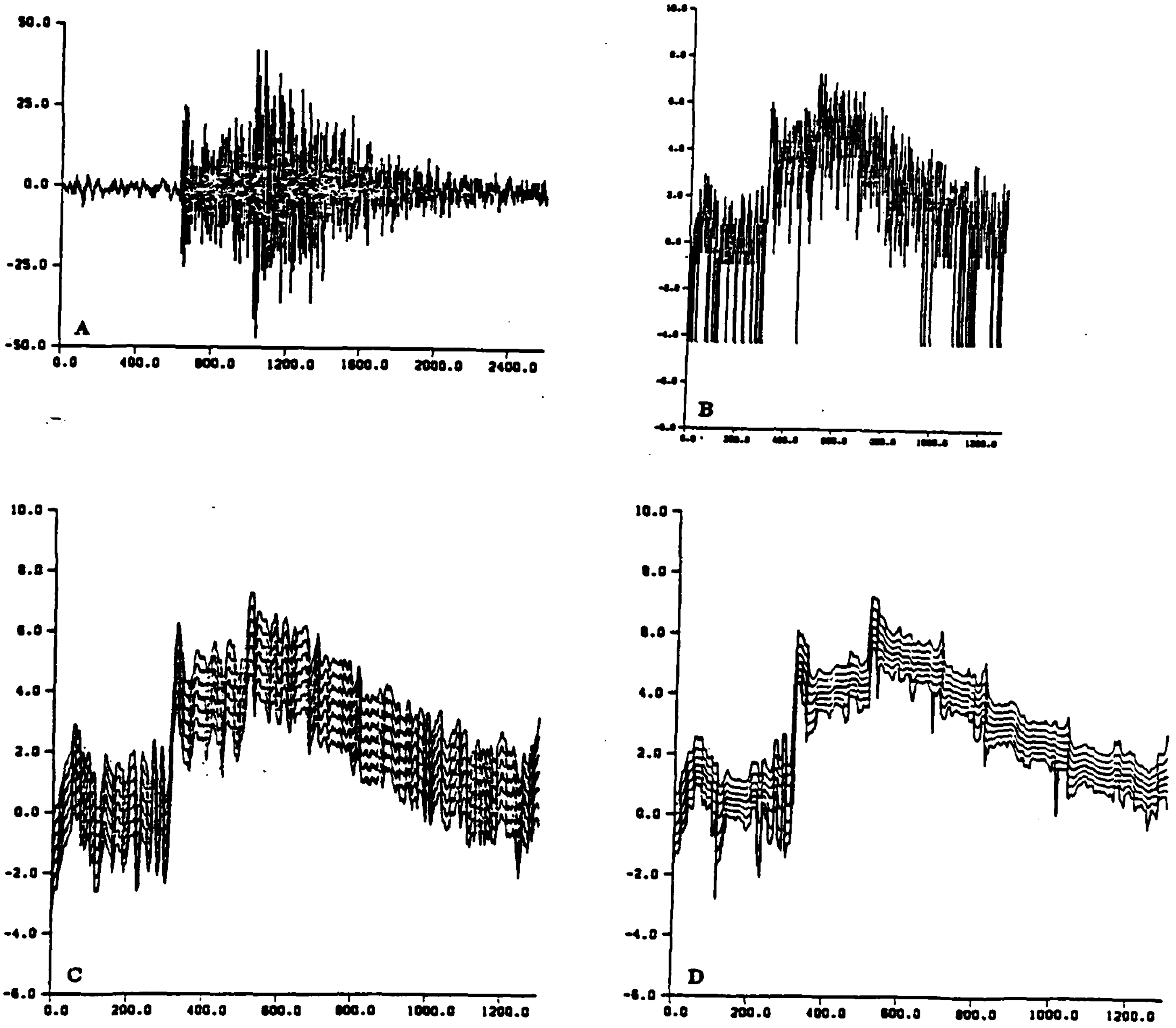

A
4
2
0
-2
-4
-6
0
100
200
300
400
B
C
D
3
2
1
0
-5
0
5
E
F
G
4.0
2.0
0.0
-2.0
-4.0
0.0
100.0
200.0
300.0
400.0

FIGURE 6


50.0
25.0
0.0
-25.0
-50.0
A
0.0
400.0
800.0
1200.0
1600.0
2000.0
2400.0
10.0
4.0
2.0
0.0
-4.0
B
0.0
200.0
400.0
1000.0
1200.0
10.0
8.0
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0
C
0.0
200.0
400.0
600.0
800.0
1000.0
1200.0
10.0
8.0
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0
D
0.0
200.0
400.0
600.0
800.0
1000.0
1200.0

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 391	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Smoothness Priors In Time Series		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Will Gersch and Genshiro Kitagawa		8. CONTRACT OR GRANT NUMBER(s) N00014-86-K-0156 and N00014-83-K-0238
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 411SP		12. REPORT DATE June 2, 1987
		13. NUMBER OF PAGES 55
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Bayesian model, smoothness priors, non Gaussian time series, stationary time series, nonstationary time series.		
20 ABSTRACT (Continue on reverse side if necessary and identify by block number) PLEASE SEE FOLLOWING PAGE.		

TECHNICAL REPORT NO. 391

20. ABSTRACT

A variety of time series signal extraction/smoothing problems are considered from a Bayesian "smoothness priors" point of view. The origin of the subject is a smoothing problem posed by Whittaker (1923). Using a stochastic regression-linear model-Gaussian disturbances framework, we model stationary time series and nonstationary mean and nonstationary covariance time series. Smoothness priors distributions on the model parameters are expressed either in terms of time domain stochastic difference equation or frequency domain constaints. A small number of (hyper)parameters specify very complex time series behavior. The critical computation is the likelihood of the Bayesian model. Finally we show a smoothness priors state space - not necessarily Gaussian - not necessarily linear model of nonstationary time series.

END
8-87
DTIC